

Center for American Progress



FUTURE OF TESTING IN EDUCATION

Effective and Equitable Assessment Systems

By Laura Jimenez and Jamil Modaffari September 2021

Contents

- 1 Introduction and summary**

- 3 What are standardized assessments and what purpose do they serve?**

- 5 History of standardized assessments in the United States**

- 10 Is the issue with the test, or how the test is administered and used?**

- 18 Recommendations: The role that assessments should play in education**

- 22 Conclusion**

- 23 About the authors and acknowledgments**

- 24 Endnotes**

Introduction and summary

[This series](#) is about the future of testing in America's schools. Part one of the series—this report—presents a theory of action that assessments should play in schools. Part two reviews advancements in technology, with a focus on artificial intelligence that can powerfully drive learning in real time. And the third part looks at assessment designs that can improve large-scale standardized tests.

Assessments are a way for stakeholders in education to understand what students know and can do. They can take many forms, including but not limited to paper and pencil or computer-adaptive formats. However, assessments do not have to be tests in the traditional sense at all; rather, they can be carried out through teacher observations of students or portfolios of students' work. Regardless of form, when assessments are well designed and a component of a system of teaching and learning that includes high-quality instruction and materials, they are part of the solution and not a source of the problem. Thus, debates on whether or not to assess students fail to create a worthwhile discussion about testing in schools and how to make assessments better.

When they are well built, standardized and nonstandardized assessments play a useful role in providing educational equity—that is, helping all students achieve at high levels. Accordingly, this report offers an alternative to the argument that all assessments are harmful: an idea for what role all assessments should play in education and the federal and state policy structure needed to make this a reality.

Assessments—in particular, one annual standardized assessment of all public school students in reading and math—became the law of the land starting in 2001 with the renewal and renaming of the Elementary and Secondary Education Act of 1965 as the No Child Left Behind Act. The rationale for this policy is to promote equity in educational opportunity by measuring how well the public education system teaches students to master a state's academic standards in these subjects.

Despite this laudable goal, federally required assessments are at times criticized because America's students have made little progress since 2001 and their results correlate with race and socioeconomic status. However, the reality remains that one assessment alone is insufficient to solve the problem of inequity in education. That is because state standardized assessments look back at the end of the year and evaluate whether students learned the state's academic standards in reading and math. They are not designed to provide information to guide teachers' daily interactions with students. This type of high-quality information, as well as professional development in how to use student data effectively, is needed to drive learning forward. Thus, the state assessment must be part of a broader system of assessments that produce data that can evaluate, inform, and predict learning to help achieve educational equity.

The Center for American Progress found that while some of the criticisms of assessments—the annual state standardized assessment in particular—are valid and must be addressed, not all of them have merit. Too often, the criticisms tend to suggest that all standardized testing is harmful or not useful.

Still, improvement at the national scale is needed. The level of research and innovation required to make assessments effective can only be achieved through federal investment and programs. For example, federal policies should invest more in assessment development, through the pilot authorized in the Every Student Succeeds Act (ESSA) and through existing assessment funding programs in that law. The federal government can play an important role in researching assessment design and understanding where and how it has a disparate impact on students. Finally, the federal government should use its resources to help ensure that teachers and school leaders become masters of local assessment development and use so that they have the data they need—when they need it—to guide student learning. And for their part, state policies should develop well-thought-out systems of assessments that are based on the state's learning standards and curriculum.

What are standardized assessments and what purpose do they serve?

Large-scale standardized assessments are just one type of assessment used in schools, with two main purposes: The first is used to predict student performance against a set of benchmarks, while the second is used to understand how many of the test's benchmarks students reached at the end of the year. That is, standardized assessments can be designed to be predictive or evaluative.

A standardized assessment presents test-takers with the same questions or the same types of questions and is administered and scored in the same way.¹ Designed to provide consistent results, standardized tests allow for comparisons between students in a single year and over time.

Standardized assessments play a prominent role in education in the United States; all public school students must take assessments in reading and math each year in grades three through eight as well as once in high school.² These tests measure what students know and can do against common state-developed, grade-level standards.

What is an academic standard?

An academic standard is what a student should know and be able to do in a particular subject.³ For example, a second grade student in math should know the ones and tens places in double-digit numbers.

Teachers are required to teach these standards and can use a wide variety of instructional materials and approaches to guide their students' learning. Thus, while federal law requires states to have the same academic standards for all students within the state, given the different levels of quality of instructional materials and practices, not all students have the same opportunity to learn and master those standards.⁴

The other types of assessments in education, which are used to predict student performance and inform instruction, are used more frequently during the school year and help to guide teachers, administrators, and even parents in providing students with the right supports at the right time.⁵ Importantly, not all assessments produce a numbered score. Some, for example, can take the form of teacher observations of student work and produce a descriptive assessment.

3 technical qualities of assessments required by federal law

This text box provides definitions of validity, reliability, and comparability in assessments and why they matter:⁶

- **Validity** refers to how accurately and fully a test measures the skills it intends to measure.⁷ For example, if an algebra test includes some geometry questions, that test is not a valid measure of algebra.
- **Reliability** refers to the consistency of the test scores across different testing sessions, different editions of the test, and when different people score the exam. Reliability indicates how consistently the test measures the knowledge and skills it should as well ensures that it is not measuring error.⁸
- **Comparability** allows for the comparison of test scores even if students took the test at different times, in different places, and under different conditions.⁹ For example, test developers will design a test that may be administered via computer or via paper and pencil to account for these differences so results can be compared.

These requirements help ensure apples-to-apples comparisons between the results of the test given on different days and under different conditions.

A state assessment's technical qualities are one tool to prioritize equity because they help ensure test results can answer the question, "How well are all students meeting a state's college- and career-ready standards in reading and math and growing in their knowledge?" Here, "college- and career-ready" means that when students meet or exceed the academic standards in reading and math, they qualify to enroll in credit-bearing courses in college. They do not need to take remedial classes to make up for unfinished learning needed for credit-bearing courses.

History of standardized assessments in the United States

Schools and standardized testing began in the first 100 years of the United States' founding, and it did so against a backdrop of systemic racism and white supremacy.¹⁰ Prior to the 1860s and emancipation, enslaved people and nonwhite people were barred from accessing education and were often punished if it was discovered they were learning to read and write. Education during this time period was reserved for the white elite.¹¹

It was within this context that the first uses of standardized tests in American education began.

College admissions, general intelligence, and K-12 achievement tests

Like in education in America, there is a deep history of racism within standardized assessments. The earliest use of assessments in America were oral qualifying exams for college admissions prior to 1840.¹²

That year marked the first uses of standardized tests in public schools. In 1845, educator Horace Mann developed common exams for school students in Boston Public Schools in an attempt to understand the quality of teaching and learning.¹³

Mann's test sparked psychologist Edward L. Thorndike's quest for other measures of intelligence, believing that society would benefit from systematic sorting and segregation of students by academic ability. Seven states—California, Kansas, Massachusetts, Michigan, New Jersey, New York, and Pennsylvania—used these exams from 1900 to 1910.¹⁴

In 1905, commissioned by the French government, psychologist Alfred Binet developed an intelligence test to identify learning deficiencies, describing “slow children who would not profit significantly from schooling.”¹⁵ And in 1916, Stanford University psychologist Lewis Terman took Binet's original test and created the

Stanford-Binet Intelligence Scales to sort students by ability into college or vocational pathways. This exam is still used today.¹⁶

The U.S. Army used a multiple-choice test to measure soldiers' mental abilities for the purpose of sorting, assigning, and discharging them during World War I. This test would become the model for future standardized assessments.¹⁷ In 1919, Terman transformed the Army Alpha test into the National Intelligence Tests for school students, selling more than 400,000 copies in 11 months.¹⁸

Testing to uphold white supremacy

Literacy tests were used to disenfranchise Black men after ratification of the 15th Amendment in 1870, until enforcement of the Voting Rights Act of 1965 outlawed literacy tests and other methods of keeping Black people from exercising their right to vote.¹⁹

Throughout the beginning of the 20th century, intelligence tests were used to determine which immigrants were “undesireable” and should not be permitted entry into the country.²⁰ Federal law in 1915 required anyone who failed the test to be turned away.²¹

Testing for public school evaluation and accountability

The use of standardized testing in schools spread nationwide after Iowa developed tests for its high school students. In 1935, the first Iowa Test of Basic Skills was administered to students in grades six through eight. Other states began using the Iowa assessment, which remained the most-used achievement test in the nation for 50 years.

But the role of testing in schools shifted in the 1970s after the then-U.S. commissioner of education created the first National Assessment of Educational Progress (NAEP) in 1969.²² It sought to provide a snapshot of the progress of education in America, using the latest testing technology to produce sound and reliable results by assessing a representative sample of the nation's students.

The NAEP marks the modern era of standardized assessments in schools to evaluate learning and a shift away from measuring intelligence toward measuring academic standards.

In the 1990s, a handful of states developed statewide testing systems that used sampling methods much like the NAEP to take a representative snapshot of how well students perform on academic tests in various subjects.²³

Federal K-12 education laws and testing

Simply put, today's federal K-12 education laws ask states, in exchange for federal funding, to ensure that students are meeting grade-level benchmarks in reading and math. These benchmarks are set by experts in math and reading as well as in psychometrics, or the measurement of learning.²⁴ The benchmarks increase in complexity grade by grade, so when students complete their education in 12th grade, they are ready for the academic demands of college—whatever form that might take—or their chosen career path. This report and the law refer to standards like these as high or rigorous standards. Federal law also asks states to evaluate schools based on these results and report these results publicly.

Federal policy did not start out this way; it has evolved since the 1990s, when states began adopting academic standards. The federal Improving America's Schools Act of 1994 asked states to apply the same standards in reading and math to all students and to assess their progress in learning the standards for the first time.²⁵

In 2001, Congress updated the Improving America's Schools Act, renaming it as the No Child Left Behind Act. The updated law required states to use those test results to evaluate schools and identify which ones needed improvement.²⁶ States published those results publicly every year and gave them to parents. A 2011 federal initiative under the Obama administration, called ESEA Flexibility, allowed states to use additional criteria to evaluate schools, but most states' criteria primarily consisted of standardized test scores.²⁷

The follow-up to the No Child Left Behind Act, now called the Every Student Succeeds Act, maintains much of the policy from ESEA Flexibility. For example, ESSA asks states to use both test scores and other criteria to evaluate all public schools. It also requires states to identify a subset of its lowest-performing schools for which to provide additional support to help them improve.

The debate about standardized testing in schools often tends to miss that the assessment requirements in federal law serve a purpose: They are one way the law helps to ensure all students receive a high-quality education through the public education system. At its heart, ESSA is a civil rights law, providing additional resources

to low-income students. It also protects the quality of education by asking states to ensure that all children learn the knowledge and skills that will help them in college and their careers. Measuring student progress toward a state's learning standards through an annual assessment is one way to know whether all students are on track or not, using a common measuring stick.

The civil rights goals of federal K-12 education laws

The first version of ESSA, then called the Elementary and Secondary Education Act (ESEA), was built on the heels of *Brown v. Board of Education* in 1954 and the Civil Rights Act of 1964, which both aimed to tackle segregation and discrimination. ESEA intended to give students from families with low resources a chance at equal education.²⁸ Likewise, the Individuals with Disabilities Education Act of 1975 (IDEA) ensured that all students with disabilities received this same opportunity.²⁹ Prior to the passage of IDEA, many students with disabilities were excluded from traditional classrooms.³⁰

The first version of ESSA centered on the role that money plays in education. It gave additional funding to schools in under-resourced communities whose local property tax bases did not provide the same amount of resources as schools in wealthy neighborhoods.

ESSA's role more recently evolved to address not just education funding but also its quality. After years of flatlining, and even declining, results in educational outcomes, documented by the U.S. Department of Education report "A Nation at Risk,"³¹ the law began to play a role in the effort to improve the quality of public school education.³² At the heart of this law—and what it requires—is an effort to include all students in public education and to hold them to the same high expectations. It does so by asking states to use their annual assessments as one measure of educational progress.

The goals of inclusion, high standards, and educational progress are the right goals. And the law has been effective in getting states to raise their standards and to include in states' accountability mechanisms for schools all students' progress on those standards. For example, states must certify that students who meet the standards when they graduate high school can enroll in college without needing remedial coursework to catch up on missed learning. And every year, states must calculate how many students met grade-level benchmarks against the goal of 95 percent of students, or the actual number of students who took the state assessment if it is lower than 95 percent.

But incremental and disproportionate progress on test scores between student groups suggests the law has been less effective in ensuring educational progress. This is in part because education is a complex process where students learn information, gain experience, and make sense of it all in a way that is useful to forging one's path in life.³³ The complexity of learning at any age cannot and should not primarily be measured by a single test.

Furthermore, research documents that students' basic needs must be met for them to be ready to learn. This is especially true for the young mind in childhood, during which it develops more than at any other time in life.³⁴

However, as the adage goes, you cannot manage what you cannot measure—meaning that while state assessments are not the silver bullet to improving the education system, they are a critical part of that process. The role that state standardized assessments should play in education is to improve the teaching and learning system. State test results at the school district level, for example, should inform what resources and supports teachers need to improve their instruction. At the state level, results can inform state efforts to provide more resources to districts needing additional supports.

Is the issue with the test, or how the test is administered and used?

The opposition to using standardized testing in schools, in part due to the history of racism in the tests, is understandably not just a historical phenomenon; vestiges of this past remain in today’s tests. Racism in testing is something that needs to be unpacked and addressed fully.

That said, some critics of standardized testing in schools miss that there are distinct issues to be acknowledged and then addressed—issues that reside with the test itself, how the test is given, and how the test results are used. These issues are too frequently treated as if they are one, so critics’ response is often to throw out the annual state test entirely. But to address these issues, and to make future assessments of student learning better, policymakers must understand these issues distinctly, as each will require different policy remedies and technical fixes.

Criticisms of annual state standardized tests include: The tests are biased; they take too long to complete; students experience stereotype threat, which is an unconscious response to a negative stereotype about a certain group by a member of that group,³⁵ when taking tests; the results are not useful for teachers; the use of these tests narrowed the curriculum used in schools; they resulted in teaching to the test; and the results are used to take money away from schools.

The authors organize these criticisms as outlined in Table 1.

TABLE 1
An organizing framework highlighting common criticisms of state standardized assessments

	Punishes students	Punishes teachers	Punishes schools
Assessment itself	Assessment is biased	Do not know what students will be tested on	N/A
How the assessment is given	Assessment is too long	Teachers are incentivized to teach to the test; narrows curriculum	Hijacks the computer lab
How the assessment results are used	Causes stereotype threat or reinforces a sense of failure	Results are not useful for teachers; teachers can get fired	Schools get less money based on results

Source: Created by the authors.

Not all criticisms of the state assessment have merit, nor do their real or perceived impacts affect students, teachers, and schools in the same way. To illustrate this, Table 1 presents these criticisms according to their real or perceived impact as well as by whether the impact is based on the assessment itself, how the assessment is administered, or how the assessment results are used. Understanding the effects and their source will help policymakers, administrators, and educators identify appropriate solutions for the root cause.

This section dissects these criticisms using a fact-based review to examine the purported impacts and deem whether they have merit and to what degree. The analysis cites claims where necessary.

Common criticism: The state standardized test is biased

Is this claim true or false? It is true, but there are multiple issues to understand.

When it comes to assessments, the term “bias” has a specific meaning. Bias happens when student inputs (their answers) are misinterpreted, misevaluated, and then scored differently.

There are three areas of the test where bias might occur: what it is trying to measure (the construct), how it is trying to measure (the method), or the test question itself (the item):

- Construct bias is an error in the measurement of the skill; for example, an item is looking to measure verbal skills but instead measures listening skills.
- Method bias is an error in the sample of students, the test form itself (in which the form is confusing), or the administration (how the test forms are given to students and collected is confusing).
- Item bias is where the test question itself is ambiguous or can result in a low or high familiarity among certain test-takers due to cultural influence.

Since standardized testing began, there have always been racial patterns in the results, suggesting bias. While this bias was initially by design, as outlined in the history of racism and assessments section of this report, modern assessment development techniques seek to eliminate bias. However, a 2010 study of the SAT confirmed a particular type of bias within its assessments: item bias.

Bias in the news

A study of the SAT published in 2010 found that harder test items favored Black test-takers and easier questions favored white.³⁶ That is, Black students taking the test more often answered the harder questions correctly at a higher rate than white students, who more often answered the easier questions correctly.

Why would this situation be the case?

Researchers theorize that the easier questions use more casual, everyday language that is part of white dominant culture. The study also suggests that the way the SAT is scored holds Black student scores down because the easier questions receive more points.

Therefore, if the harder questions received more points, this bias could be addressed when the test is scored. This example shows that even if there is bias in any aspect of the test, that bias can be balanced out in the scoring process. However, this method is not a complete answer to rooting out bias or its impact in assessments.

When there are consistently racial patterns in the results, bias is present somewhere in the test, whether in the construct, method, or item, or some combination of these.

Bias in state standardized tests for school accountability

Regarding state standardized tests used for federal accountability, states undergo rigorous analysis to detect and remove all types of bias and must submit this evidence to the U.S. Department of Education for their assessment systems to be approved.

The department runs a peer review process of each state's assessment system, with experts in test development and curriculum as well as teachers and local assessment administrators. The data are in the form of analysis of test-taker results; it is not possible to analyze a test before it is given to test-takers because experts must see evidence of how the students interacted with the test.

There is no question that bias exists in standardized testing, and there is no question that standardized testing is one tool to understand how students are doing against common, challenging standards of learning. Therefore, future versions of tests should pilot new test items on a broader range of students from different racial and ethnic backgrounds to minimize—or better yet eliminate—bias. And where bias occurs in test construct, how the skills and knowledge are being assessed must be better understood and addressed.

However, while eliminating bias is a necessary step, it is not sufficient on its own if educators are to take an anti-racist approach to teaching and learning.

Assessments and cultural competence

Solely addressing bias in assessments is an incomplete response; assessments are one part of a larger system of teaching and learning, which also includes standards, instructional materials, and instructional practice. These elements must be evaluated for bias and addressed as an entire system.

The question of synchronizing these parts of the teaching and learning experience extends beyond the academic content, to the interactions between teachers and students, the climate of the school, and the larger sociocultural context. This entire process is known as cultural synchronization, and the practice of applying it to teaching is called culturally responsive pedagogy.³⁷

Culturally responsive pedagogy is an attempt to create continuity between what students experience at home and what they experience at schools. Students see people who share their race and ethnicity included in the topics they study, for example. When students have access to culturally responsive pedagogy, they experience academic success and develop cultural competence and the ability to question the current social order, thereby allowing students to gain a sense of empowerment from their history and culture.³⁸

Common criticism: It is hard for teachers to prepare students without knowing the test material

Is this claim true or false? It is neither.

Federal law places a premium on test security. Test results are used to evaluate school performance and must be a fair and accurate representation of student participation in the test. Therefore, only test developers know the exact test items.

This does not mean that teachers do not know what kinds of test items will be included, however. Two groups—the Smarter Balanced Assessment Consortium (SBAC) and the Partnership for Assessment of Readiness for College and Careers (PARCC)—developed an annual statewide assessment that more than 40 states first used in 2014. Each group releases test items used on previous tests, which reflect the kinds of items included in future tests.³⁹ Additionally, both consortia analyze student responses as a resource for teachers.

Common criticism: The tests take too long, lessening time for instruction

Is this claim true or false? It is neither, because the answer depends on what consumers of the test results value knowing about student performance.

The SBAC and PARCC assessments take about eight to nine hours total to complete for reading and math. This lengthy assessment is due to the law requiring that the test measures the full range and depth of the state’s grade-level standards through formats that are not just multiple choice but also constructed response, or written answers. These types of test questions take more time to complete than filling in an answer bubble.

If policymakers, educators, and parents value knowing how well students have learned the full set of standards for their grade level—and not just responses to “yes” or “no” multiple-choice answers—then an eight- or nine-hour test will provide that information.

However, if policymakers, educators, and parents prefer having a high-level overview of whether students have learned at grade level, then a shorter exam will suffice.

Common criticism: Students experience stereotype threat when taking the test

Is this claim true or false? It is to be determined.

Stereotype threat occurs when a member of a certain social group is at risk of experiencing an unconscious response to a negative stereotype about that person’s own group.⁴⁰

More than 300 studies show the impacts of stereotype threat, which include limiting one’s aspirations in a field of study or career. Most of the studies are on college students and other adults, and none are specific to the state test. The studies include how participants performed on tasks as well as on tests.⁴¹

Like bias, this is a phenomenon that probably exists among public school students and is an area that needs to be studied further to know how it might be affecting student performance and how it can be mitigated.

Common criticism: Results are not useful for teachers to employ in their practice

Is this claim true or false? It is false.

Annual state assessments results are used to evaluate, at year's end, whether students met the state's academic standards for their grade. As a result, they are not designed for teachers to use in their daily practice to customize instruction for students.

However, annual state assessments will show patterns for entire classrooms of students, and teachers can use this information to know generally what students did and did not learn and adjust their approach accordingly for the next school year's students.

Common criticism: State standardized tests narrowed the curriculum

Is this claim true or false? It is true, and it happened because of how the tests were used to evaluate schools.

In the period from 1987 to 2003, the amount of time devoted to different subjects held steady in public elementary schools: two hours to English, one hour to math, and a half-hour each to social studies and science.

However, since the No Child Left Behind Act passed in 2001 and required reading and math tests to be included in school accountability results, 62 percent of a nationally representative sample of schools and 75 percent of schools identified as in need of improvement increased their time given to math and English by about half and decreased the time given to other subjects.⁴²

Common criticism: The tests result in teaching to the test

Is this claim true or false? It is true.

A 2007 review of 49 studies found that 80 percent of the studies saw a change in curriculum and increased focus on teacher-led instruction.⁴³

Generally, “teaching to the test” means “teaching in a manner that is not considered optimal for learning standard content or skills, but is believed to improve test performance.”⁴⁴

Is teaching to the test always bad? Not necessarily. An assessment only covers a subset of the range of standards. If a teacher solely focuses on that subset, then students are missing out on other important content and skill development. But if tests are aligned with the depth and breadth of the standards, then teaching to the test some of the time can be advantageous.

Common criticism: Test results are used to take money from low-income schools

Is this claim true or false? It is false, although there are nuances to understand.

The No Child Left Behind Act measured school performance through a construct called adequate yearly progress (AYP). AYP referred to the total number of students who achieved a score of proficient or above on the state test. Students' scores of proficient or above were supposed to signify that they had achieved or exceeded the grade-level standards. For comparison, a grade of a C usually means proficient on an A–F grading scale.

When schools missed AYP targets, districts set aside 20 percent of their Title I funding—which provides additional education programs in schools in low-income communities—to pay for supplemental education services. These services include tutoring as well as transportation for students to attend higher-performing schools that also received Title I money.

As a result, districts did not exactly lose Title I money; however, they had to use it for a specific purpose when they did not meet AYP targets. Because there was less money to spend on services and resources for individual schools, the claim that test results lead to less money for schools does have some merit, but it is not the entire picture.

The current version of the law, ESSA, takes a different approach to deploying resources to schools that need them. It eliminates AYP and instead requires states to provide additional money to some schools that are classified as low-performing, many of which have been poorly funded for decades, to help them improve.⁴⁵

Common criticism: Test results are used to close schools serving Black and Latinx students

Is this claim true or false? It is true, but only for a subset of schools that were closed.

Between 2003 and 2013, approximately 2 percent of all U.S. public schools closed.⁴⁶ Many of these were due to declining populations. Some of them, however, were due to poor student outcomes, as indicated in part by state assessment results.

A 2017 study of closures found that 1,522 schools closed between 2006-07 and 2012-13 because their state assessment scores were in the bottom 20 percent of the 26 states included in the study. In the study, schools with higher rates of Black and Hispanic students were more likely to close than similarly performing schools with smaller shares of students who are racial and ethnic minorities.⁴⁷

Recommendations: The role that assessments should play in education

Assessments should drive excellent teaching and ensure that all students learn at high levels. To do so, education policy and practice must encompass a broader range of assessments so that schools have complete and effective assessment systems. This system would include predictive, informative, and evaluative assessments based on the state's standards and curriculum. Such a system would be based on three principles:

1. **Assessments should be used only for their three intended purposes:** 1) to predict student performance, 2) to inform instruction, or 3) to evaluate learning.⁴⁸ A complete system would include assessments that serve one of these purposes, and there should not be too many assessments that serve the same purpose.
2. **All assessments should align with the state's academic standards and with high-quality instructional materials.** This alignment allows for a tighter integration between what students learn in class and the items that will be included on the assessment.⁴⁹ Additionally, assessment results will send consistent signals about how well students are learning the standards and what they must continue to learn to master the full range by year's end.
3. **Effective assessment systems use the data appropriately and for the right audiences.** For example, teachers should use predictive assessment results to inform what standards students must still learn to be able to grasp the content of the first lesson in a unit of instruction. Predictive assessments can also shed light on whether students are on track to meet benchmarks on end-of-year or other evaluative assessments. On the other hand, district administrators and policymakers can use evaluative assessment results to inform what types of additional supports students may need to achieve the standards. Because of the different tools policymakers, administrators, and educators must deploy in the education system, all of these stakeholders must be informed on how to appropriately interpret assessment results.

Assessment audits can be an effective tool to guide states' and districts' understanding of what assessments they currently use and what purpose they serve, ensuring that students are not overassessed for evaluation purposes but also to give information that predicts their learning and informs instruction.⁵⁰ Conducting such an audit can be a useful first step in building an effective and balanced assessment system.

A case study for local assessments: Finland

For years, Finland enjoyed top rankings on the Program for International Student Assessment, an international test of 15-year-olds in Organization for Economic Cooperation and Development member countries. Though Finland consistently outranks most other countries, its scores have declined—a puzzling development for researchers.⁵¹

Despite the decline, Finland credits its educational success to its heavy investment in teacher training as well as its model for using independent and group projects as ways to engage students in their learning.

Notably, Finland takes a very different approach to assessments than the United States. The country eliminated its national evaluative assessment and instead allows teachers to design their own assessments that are based on the national curriculum.⁵² The same training equips teachers to design school-based projects for students. The country only uses predictive standardized assessments for students to take at the end of their education. Those results are used for consideration for college admissions, not for evaluating education programs, students, or teachers.

As a comparison in the United States, the quality of teacher development and training varies widely and does not uniformly reflect the quality of the opportunities available in Finland or other high-achieving countries.⁵³

While Finland's example showcases the power of assessments to inform instruction, that is only one aspect of a complete and effective assessment system.

How federal policies can support the effective use of assessment in teaching and learning

Effective and complete assessment systems contribute to student learning.⁵⁴

Currently, however, states are not required to have such systems in place—only an annual evaluative assessment, which is a practice that began with the 2001 No Child Left Behind Act. As a result, the focus on one test and its use to evaluate schools created bad incentives, as discussed earlier in this report. However, that does not mean that evaluative tests should not play a role in education; to the contrary, high-quality assessments aligned to a state’s standards are a critical tool in the teaching and learning process. Federal policy should thus push states and districts to establish and maintain effective and balanced assessment systems.

Accordingly, any future updates to ESSA should ask states to design a vision for how assessments are part of the teaching and learning process and then describe which assessments predict student performance, inform teaching and learning, and evaluate student learning.

At same time, more large-scale research and development is needed in the practice and use of assessments. This is a great role for the federal government to play. The federal government should:

- Fund the assessment pilot in ESSA and loosen some of its restrictions to support states in trying more innovative designs, even if those designs do not pan out. The assessment pilot gives states five years to try new assessment designs to replace the state’s annual evaluative assessment.
- Fund the development of new ways to assess students across a broad range of skills and not just through tests, but through other demonstrations of student learning as well. This can be done outside of the assessment pilot through the Competitive Grants for State Assessments program.
- Fund the development of predictive, informative, and evaluative assessments.
- Fund the creation of new and better ways to report assessment results, not just to parents, but to teachers and policymakers as well.
- Study bias in testing, including in testing construct, methods, and items.
- Reshape the teacher-focused Title II of ESSA and the Higher Education Act of 1965 so that they promote the development of better teacher training and support when it comes to assessment use.

- Partner with institutes of higher education to identify ways that the training of future psychometricians—scientists who develop assessments—should change to ensure that tomorrow’s tests do not replicate the bias and other drawbacks seen in current versions.
- Encourage states to create science, technology, engineering, and math (STEM) pathways that expose students to future careers as psychometric measurement experts to create a more diverse pool of psychometricians.

Conclusion

Today's conversation around assessments is dysfunctional and a zero-sum game. Assessments are neither useless nor are they the silver bullet to improving education. Instead, they are a vital tool to drive excellent teaching and learning. For equitable and effective testing to be fully realized, policymakers must invest in understanding the limits of today's assessments and build on those lessons. Innovation and research should support the development of assessment systems that drive teaching and learning forward as well as evaluate student learning.

The society and workforce of tomorrow will require students not only to master academic skills but also possess a broad range of crosscutting knowledge and abilities. America's current assessment system does a poor job of measuring how well students are prepared for that future and of guiding educators and parents to support students in their development. That is what policymakers and educators must address as they consider the future of assessments.

To that end, future research by the Center for American Progress will highlight ways in which technology advances may support the measurement of a broader range of student knowledge and skills.

About the authors

Laura Jimenez is the director of standards and accountability on the K-12 Education team at the Center for American Progress.

Jamil Modaffari is a research assistant for K-12 Education at the Center.

Acknowledgments

The authors would like to thank the following people for their advice in writing this report:

- Abby Javurek, Northwest Evaluation Association
- Alina von Davier, Duolingo
- Ashley Eden, New Meridian
- Bethany Little, EducationCounsel
- Edward Metz, Institute of Education Sciences
- Elda Garcia, National Association of Testing Professionals
- Jack Buckley, Robox; American Institutes for Research
- James Pellegrino, University of Illinois at Chicago
- John Whitmer, Institute of Education Sciences
- Krasimir Staykov, Student Voice
- Kristopher John, New Meridian
- Laura Slover, CenterPoint Education Solutions
- Margaret Hor, CenterPoint Education Solutions
- Mark DeLoura, Games and Learning Inc.
- Mark Jutabha, WestEd
- Michael Rothman, independent consultant; formerly of Eskolta School Research and Design
- Michael Watson, New Classrooms
- Mohan Sivaloganathan, Our Turn
- Neil Heffernan, Worcester Polytechnic Institute
- Osonde Osoba, RAND Corp.
- Roxanne Garcia, UnidosUS
- Sandi Jacobs, EducationCounsel
- Sean Worley, EducationCounsel
- Scott Palmer, EducationCounsel
- Terra Wallin, The Education Trust
- Tim Langan, National Parents Union
- Vivett Dukes, National Parents Union

Endnotes

- 1 Great Schools Partnership Glossary of Education Reform, "Standardized Test," available at <https://www.edglossary.org/standardized-test/> (last accessed May 2021).
- 2 Every Student Succeeds Act, Public Law 114-95, 114th Cong., 1st sess. (December 10, 2015), available at <https://www.govinfo.gov/content/pkg/PLAW-114publ95/pdf/PLAW-114publ95.pdf>.
- 3 U.S. Department of Education, "A State's Guide to the U.S. Department of Education's Assessment Peer Review Process" (Washington: 2018), available at https://www2.ed.gov/admins/lead/account/saa/assessmentpeer-review.pdf?utm_content=&utm_medium=email&utm_name=&utm_source=govdelivery&utm_term=.
- 4 Lisette Partelow and Sarah Shapiro, "Curriculum Reform in the Nation's Largest School Districts" (Washington: Center for American Progress, 2018), available at <https://www.americanprogress.org/issues/education-k-12/reports/2018/08/29/454705/curriculum-reform-nations-largest-school-districts/>.
- 5 Achievement Network, "Teaching Comes First: How School District Leaders Can Support Teachers, Save Time, and Serve Students with a New Vision for Assessment" (Boston: 2018), available at https://www.flipsnack.com/ANetOrg/anet_whitepaper_2018_web/full-view.html?mkt_tok=eyJjoiWVdNNU5TWTZlbnVfZjNjYndylnQjOjIjOGRLEZMKzcxelhkVzZSIIUUpza1wvTERkakt1YzRWNvVFVlRrR1NhbzdSR256cVJWt3lpZnM0dThud0N4VTJONlh1aUM5Z-W5obllbVBUCGY4ZFZHUG9yMXU1c25jM05lX9sZE5KZV-ROQ0hhcVdKdE4MXByN1RKOGZVRXdaln0%3D.
- 6 U.S. Department of Education, "Standards, Assessments and Accountability," available at <https://www2.ed.gov/admins/lead/account/saa.html> (last accessed May 2021).
- 7 Fiona Middleton, "The four types of validity," Scribbr, September 6, 2019, available at <https://www.scribbr.com/methodology/types-of-validity/>.
- 8 Samuel A. Livingston, "Test Reliability—Basic Concepts" (Princeton, NJ: Educational Testing Service, 2018), available at <https://www.ets.org/Media/Research/pdf/RM-18-01.pdf>; U.S. Department of Education Institute of Education Sciences, "Assessing the Reliability of State Assessments," November 2009, available at https://ies.ed.gov/ncee/pubs/2009013/section_2b.asp.
- 9 Amy I. Berman, Edward H. Haertel, and James W. Pellegrino, "Comparability of Large-Scale Educational Assessments: Issues and Recommendations" (Washington: National Academy of Education, 2020), available at <https://naeducation.org/wp-content/uploads/2020/06/Comparability-of-Large-Scale-Educational-Assessments.pdf>.
- 10 Lehigh University College of Education, "History of Standardized Testing," October 18, 2013, available at <https://ed.lehigh.edu/news-events/news/history-standardized-testing>; PBS Frontline, "History of the SAT: A Timeline," available at <https://www.pbs.org/wgbh/pages/frontline/shows/sats/where/timeline.html> (last accessed July 2021); Iowa PBS, "Standardized Testing Begins in Iowa," available at <https://www.iowapbs.org/iowapathways/mypath/all-those-tests> (last accessed July 2021); U.S. Department of Education National Center for Education Statistics, "History and Innovation," available at <https://nces.ed.gov/nationsreportcard/about/timeline.aspx> (last accessed July 2021).
- 11 Cornell University Library, "In Their Own Words: Slave Narratives," available at <https://rmc.library.cornell.edu/abolitionism/narratives.htm#:~:text=For%20slaves%20and%20their%20teachers,find%2C%20imprisoned%2C%20or%20whipped> (last accessed May 2021); Smithsonian American Art Museum, "Literacy as Freedom," available at <https://americanexperience.si.edu/wp-content/uploads/2014/09/Literacy-as-Freedom.pdf> (last accessed May 2021); Robert Greene II, "Poor Whites Have Been Written out of History for a Very Political Reason: An Interview With Keri Leigh Merritt," *Jacobin*, August 24, 2019, available at <https://jacobinmag.com/2019/08/poor-whites-have-been-written-out-of-history-for-a-very-political-reason>.
- 12 Erik the Red, "A (Mostly) Brief History Of The SAT And ACT Tests," available at <https://www.erikthered.com/tutor/sat-act-history-printable.html> (last accessed May 2021).
- 13 Carole J. Gallagher, "Reconciling a Tradition of Testing with a New Learning Paradigm," *Educational Psychology Review* 15 (1) (2003): 83–99, available at https://www.jstor.org/stable/23361535?mag=short-history-standardized-tests&seq=4#metadata_info_tab_contents.
- 14 Ibid.
- 15 Ibid.
- 16 Stanford-Binet Test, "Home," available at <https://stanfordbinettest.com/> (last accessed May 2021).
- 17 Carole Sanger Brink, "A Historical Perspective of Testing and Assessment Including the Impact of Summative and Formative Assessment on Student Achievement" (Hattiesburg, MS: University of Southern Mississippi, 2011), available at <https://aquila.usm.edu/cgi/viewcontent.cgi?article=1438&context=dissertations>.
- 18 Gallagher, "Reconciling a Tradition of Testing with a New Learning Paradigm."
- 19 University of Michigan, "Techniques of Direct Disenfranchisement, 1880-1965," available at <http://www.umich.edu/~lawrace/disenfranchise1.htm> (last accessed May 2021).
- 20 Lorraine Boissoneault, "Literacy Tests and Asian Exclusion Were the Hallmarks of the 1917 Immigration Act," *Smithsonian* magazine, February 6, 2017, available at <https://www.smithsonianmag.com/history/how-america-grappled-immigration-100-years-ago-180962058/>.
- 21 Dan Schlenoff, "Challenging the Immigrant: The Ellis Island intelligence tests, 1915," *Scientific American*, January 1, 2015, available at <https://www.scientificamerican.com/article/ellis-island-challenging-the-immigrant/>.
- 22 U.S. Department of Education National Center for Education Statistics, "History and Innovation."
- 23 Margaret E. Goertz and Mark C. Duffy, "Assessment and Accountability Systems in the 50 States: 1999-2000," *CPRE Research Reports* (2001), available at https://repository.upenn.edu/cgi/viewcontent.cgi?article=1068&context=cpre_researchreports.
- 24 TeAchnology, "Who creates Learning Standards?," available at <http://www.teach-nology.com/teachers/standards/who.html> (last accessed June 2021).

- 25 U.S. Department of Education, "The Improving America's Schools Act of 1994: Reauthorization of the Elementary and Secondary Education Act," available at <https://www2.ed.gov/offices/OESE/archives/legislation/ESEA/brochure/iasa-bro.html> (last accessed May 2021).
- 26 U.S. Department of Education, "No Child Left Behind: A Desktop Reference" (Washington: 2002), available at <https://www2.ed.gov/admins/lead/account/nclbreference/reference.pdf>.
- 27 U.S. Department of Education, "ESEA Flexibility," available at <https://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html> (last accessed June 2021).
- 28 Elementary and Secondary Education Act of 1965, Public Law 89-10, 89th Cong., 1st sess. (April 9, 1965), available at link <https://www2.ed.gov/documents/essa-act-of-1965.pdf>.
- 29 Kyrie E. Dragoo, "The Individuals with Disabilities Education Act (IDEA), Part B: Key Statutory and Regulatory Provisions" (Washington: Congressional Research Service, 2017), available at <https://fas.org/sgp/crs/misc/R41833.pdf>.
- 30 National Council on Disability, "IDEA Series: The Segregation of Students with Disabilities" (Washington: 2018), available at https://ncd.gov/sites/default/files/NCD_Segregation-SWD_508.pdf.
- 31 U.S. Department of Education, "A Nation At Risk" (Washington: 1983), available at <https://www2.ed.gov/pubs/NatAtRisk/risk.html>.
- 32 U.S. Department of Education, "No Child Left Behind: A Desktop Reference."
- 33 National Academies of Sciences, Engineering, and Medicine, "How People Learn II: Learners, Contexts, and Cultures" (Washington: 2018), available at <https://www.nationalacademies.org/news/2018/10/learning-is-a-complex-and-active-process-that-occurs-throughout-the-lifespan-new-report-says>.
- 34 Gabriel Piña and others, "Being Healthy and Ready to Learn is Linked with Socioeconomic Conditions for Preschoolers" (Bethesda, MD: Child Trends, 2020), available at <https://www.childtrends.org/publications/being-healthy-and-ready-to-learn-is-linked-with-socioeconomic-conditions-for-preschoolers>; Arizona PBS, "Early childhood brain development has lifelong impact," available at <https://azpbs.org/2017/11/early-childhood-brain-development-lifelong-impact/> (last accessed May 2021).
- 35 Kirwan Institute for the Study of Race and Ethnicity, "Standardized Testing and Stereotype Threat," March 12, 2013, available at <https://kirwaninstitute.osu.edu/article/standardized-testing-and-stereotype-threat>.
- 36 Sarwat Amin Rattani, "SAT: Does Racial Bias Exist?," *Creative Education* 7 (15) (2016): 2151–2162, available at https://www.researchgate.net/publication/308390801_SAT_Does_Racial_Bias_Exist.
- 37 Gloria Ladson-Billings, "But That's Just Good Teaching! The Case for Culturally Relevant Pedagogy," *Theory Into Practice* 34 (3) (1995): 159–165, available at https://www.researchgate.net/publication/245591801_But_That's_Just_Good_Teaching_The_Case_for_Culturally_Relevant_Pedagogy.
- 38 Ibid.
- 39 Smarter Balanced, "Sample Items," available at <https://sampleitems.smarterbalanced.org/> (last accessed May 2021); New Meridian, "Released Items," available at <https://resources.newmeridiancorp.org/released-items/> (last accessed May 2021).
- 40 Steve Strossner and Catherine Good, "Stereotype Threat: An Overview" University of Arizona, available at https://diversity.arizona.edu/sites/default/files/stereotype_threat_overview.pdf (last accessed July 2021).
- 41 Ibid.
- 42 Jennifer McMurrer, "Choices, Changes, and Challenges: Curriculum and Instruction in the NCLB Era" (Washington: Center on Education Policy), available at <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=593580376CAA72933F4BB739B7C12DBD?doi=10.1.1.643.5250&rep=rep1&type=pdf>.
- 43 Ibid.
- 44 Richard P. Phelps, "Teaching to the test: A very large red herring," *Nonpartisan Education Review* 12 (1) (2016), available at <https://www.nonpartisaneducation.org/Review/Essays/v12n1.htm>.
- 45 Daarel Burnette II, "How COVID-19 Will Make Fixing America's Worst-Performing Schools Even Harder," *Education Week*, March 23, 2021, available at <https://www.edweek.org/leadership/how-covid-19-will-make-fixing-americas-worst-performing-schools-even-harder/2021/03>.
- 46 Marcus A. Winters, "Should Failing Schools Be Closed? What the Research Says" (New York: Manhattan Institute for Policy Research, 2019), available at <https://files.eric.ed.gov/fulltext/ED599013.pdf>.
- 47 Center for Research on Education Outcomes, "Lights Off: Practice and Impact of Closing Low-Performing Schools" (Stanford, CA: 2017), available at https://credo.stanford.edu/sites/g/files/sbiybj6481/fi/closure_final_executive_summary.pdf.
- 48 Achievement Network, "Teaching Comes First."
- 49 Deborah Sigman and Marie Mancuso, "Designing a Comprehensive Assessment System" (San Francisco: WestEd, 2017), available at <https://www.wested.org/wp-content/uploads/2017/03/resource-designing-a-comprehensive-assessment-system.pdf>.
- 50 Achievement Network, "Teaching Comes First"; Achieve, "Student Assessment Inventory for School Districts," available at <https://www.achieve.org/assessmentinventory> (last accessed June 2021).
- 51 Arto K. Ahonen, "Finland: Success Through Equity—The Trajectories in PISA Performance," in Nuno Crato, ed., *Improving a Country's Education: PISA 2018 Results in 10 Countries* (New York: Springer Publishing, 2020), available at https://link.springer.com/chapter/10.1007/978-3-030-59031-4_6; Yle Uutiset, "Time Out: What happened to Finland's education miracle?," January 16, 2020, available at https://yle.fi/uutiset/osasto/news/time_out_what_happened_to_finlands_education_miracle/11160051.
- 52 Linda Darling-Hammond and Laura McCloskey, "Assessment for Learning Around the World: What Would It Mean to Be Internationally Competitive?," *Phi Delta Kappan* 90 (4) (2008): 263–272, available at https://www.hkeaa.edu.hk/DocLibrary/SBA/HKDSE/Eng_DVD/doc/Educ_sa_international_assessment.pdf; Hannele Niemi, "Teacher Professional Development in Finland: Towards a More Holistic Approach," *Psychology, Society, and Education* 7 (3) (2015): 279–294, available at https://www.researchgate.net/publication/301225639_Teacher_Professional_Development_in_Finland_Towards_a_More_Holistic_Approach.

- 53 Aisha Asif, "Research suggests poor quality of teacher training programs in U.S. compared to other countries," *The Hechinger Report*, October 31, 2013, available at <https://hechingerreport.org/research-suggests-poor-quality-of-teacher-training-programs-in-u-s-compared-to-other-countries/>; Emily Richmond, "America's Teacher-Training Programs Aren't Good Enough," *The Atlantic*, June 18, 2013, available at <https://www.theatlantic.com/national/archive/2013/06/americas-teacher-training-programs-arent-good-enough/276993/>. Ruth Chung-Wei and others, "How Nations Invest in Teachers: High Achieving Nations Treat Their Teachers As Professionals," *Educational Leadership* 66 (5) (2009): 28–33, available at <https://edpolicy.stanford.edu/library/publications/128>.
- 54 Thomas R. Guskey, "How Classroom Assessments Improve Learning," *Educational Leadership* 60 (5) (2003): 6–11, available at <http://www.ascd.org/publications/educational-leadership/feb03/vol60/num05/How-Classroom-Assessments-Improve-Learning.aspx>.

Our Mission

The Center for American Progress is an independent, nonpartisan policy institute that is dedicated to improving the lives of all Americans, through bold, progressive ideas, as well as strong leadership and concerted action. Our aim is not just to change the conversation, but to change the country.

Our Values

As progressives, we believe America should be a land of boundless opportunity, where people can climb the ladder of economic mobility. We believe we owe it to future generations to protect the planet and promote peace and shared global prosperity.

And we believe an effective government can earn the trust of the American people, champion the common good over narrow self-interest, and harness the strength of our diversity.

Our Approach

We develop new policy ideas, challenge the media to cover the issues that truly matter, and shape the national debate. With policy teams in major issue areas, American Progress can think creatively at the cross-section of traditional boundaries to develop ideas for policymakers that lead to real change. By employing an extensive communications and outreach effort that we adapt to a rapidly changing media landscape, we move our ideas aggressively in the national policy debate.

