

POLICY DIALOGUE

Policy Dialogue: Twenty Years of Test-Based Accountability

Diane Ravitch, Denise Forte, Princess Moss and Paul Reville

Abstract

Since No Child Left Behind was signed into law, test-based accountability has become a core feature of the K-12 public education system in the United States. The approach, it would seem, is here to stay. Yet that is not to say that anything resembling a consensus has emerged. Over the past twenty years, critics have continued to raise questions about the theory of change underlying test-based accountability, and scholars have detailed a variety of unintended consequences associated with it.

If test-based accountability is both likely to persist and imperfect in its design, then it is critical to consider how its shortcomings might be addressed. In service of that aim, and in keeping with the mission of this feature, this Policy Dialogue explores future possibilities by starting, first, with a look at the past. In this particular case, participants were asked to address one simple question: “What have we learned from two decades of high-stakes testing?”

As regular readers of *HEQ* are aware, these dialogues usually feature a historian in conversation with a scholar or practitioner from the world of policy. In this case, the choice of Diane Ravitch was a natural one, particularly given the fact that she is a member of *HEQ*'s editorial board. A research professor at New York University, she is also a former assistant US secretary of education and the author of several books about measurement and accountability.

Rather than select a single interlocutor, however, the editors chose to pair her with three leaders who represent the broad range of viewpoints in the field: Denise Forte, Princess Moss, and Paul Reville. Denise Forte is the interim CEO of The Education Trust. She brings to our conversation twenty years of experience in congressional staff roles, including as the staff director for the House Committee on Education and the Workforce. Princess Moss is vice president of the National Education Association and cochair of the NEA's task force on measurement and accountability. In prior work with the NEA's Executive Committee, she helped develop the group's position on reauthorization of the Elementary and Secondary Education Act—from NCLB to the Every Student Succeeds Act. Paul Reville is the Francis Keppel Professor of Practice of Educational Policy and Administration at the Harvard Graduate School of Education and former secretary of education for the Commonwealth of Massachusetts. Nearly a decade before the passage of NCLB, he played a key role in the development of the Massachusetts Education Reform Act of 1993, which instituted standards-based accountability across the state.

HEQ Policy Dialogues are, by design, intended to promote an informal, free exchange of ideas between scholars. At the end of the exchange, we offer a list of references for readers who wish to follow up on sources relevant to the discussion.

Keywords: No Child Left Behind; accountability; standardized testing; educational assessment

Diane Ravitch: American schools have always used tests of some sort to determine whether students have learned what they were taught. The age of standardized tests began approximately a century ago, when educational psychologists developed intelligence tests to sort millions of army draftees and identify those who were officer material.

After the war, standardized tests were used in some schools to sort children into tracks, academic or vocational. In 1941, the College Board replaced its written college entry exams with a standardized test called the Scholastic Aptitude Test. Many different standardized tests were used by districts and states for the balance of the twentieth century.

In the 1960s, a dozen nations participated in the first international test of mathematics (US students placed last). About the same time, the US launched a national test called the National Assessment of Educational Progress (NAEP), which was taken by a sampling of students in states that volunteered. The results were reported by regions, and no stakes were attached for students, teachers, or schools.

The passage of George W. Bush's No Child Left Behind Act in 2001 marked a dramatic change in the role of standardized testing in American education. Bush claimed that annual testing, with rewards and punishments, had produced a "Texas miracle." The new law required every school in the nation to test every student from grades three to eight in reading and math. By 2014, every student was supposed to score "proficient," a high bar on the NAEP scale that is equivalent to a solid A, reached typically by 35 percent of students. Schools that were not on track to meet this ambitious (some would say impossible) goal were subject to a cascade of punishments, the worst of which meant firing the staff, turning the school over to private management, or closing it down.

NCLB did not produce an education miracle by 2014 (but then, neither had Texas). No district could say that 100 percent of its students had scored proficient.

Instead of questioning the theory behind NCLB—that high-stakes standardized testing would raise everyone's test scores and close achievement gaps—the Obama administration doubled down on the same test-and-punish strategy. The \$5 billion Race to the Top (RTTT) competition was built on the foundation of NCLB. Under NCLB, schools were held accountable. Under RTTT, both schools and individual teachers were held accountable for test scores. States that wanted to win money had to agree, among other things, to evaluate teachers based on the test scores of their students.

What happened? From 2010 to 2017 (the last test administration), NAEP scores were flat, and achievement gaps remained large. Test scores became the most important measure of schools. The use of test scores to measure teacher quality failed. The tests measured family income and education of families. Teachers in affluent districts got excellent results. Those teaching students with disabilities, students in poverty, and students who were English-language learners received poor results. Large-scale studies of test-based evaluation showed no gains for students and also showed that highly ranked teachers tried to avoid classes with high-needs students.

Here are the lessons of the past twenty years, in my view. Standardized testing is a measure, not an instructional tool. Threatening to punish teachers if scores go down or promising to give them extra money if scores go up is not a sensible or effective

way to improve education. Tests are designed to discriminate from best to worst, not to level the results. There will always be an upper half and a bottom half. The most advantaged students will cluster in the top half, and the least advantaged will cluster in the bottom half.

The obsessive devotion to standardized testing has narrowed the curriculum to emphasize the tested subjects of reading and math. But NAEP scores on these two subjects have not budged since 2010.

The importance placed on test scores has led to numerous cheating scandals. It has also, inevitably, encouraged teaching to the test, which used to be considered unprofessional. Students are not being tested based on what their teachers taught them, but on what test publishers think they should know.

The most significant lesson of the past twenty years is that we must reexamine our devotion to standardized testing as the ultimate measure of students, teachers, and schools. Education is so much more than a test score. Standardized testing, by its nature, privileges the already privileged. If we want better education, we must rethink how we evaluate students, teachers, and schools. And we must think anew about what education ideas we value most.

Denise Forte: I wouldn't use the term *high stakes*—that's hardly the case anymore. It's better to look at it as how to best support student learning, and we need multiple ways of measuring learning. As a parent, I want to know whether my nine-year-old can read, whether he can read and understand comprehensive texts for his age and grade. I want his teacher to know where he is and how to support him. I want his principal to know how other Black boys like him fare against all kids in the school. I want the district where he attends school to know how students in his school compare to students across the state and the nation.

The adoption of the No Child Left Behind Act twenty years ago ended the days of not knowing whether all students were receiving the support they need to achieve in the classroom. It introduced data and accountability systems to measure student and school performance, the most powerful tools we have at our disposal to ensure that all students—regardless of their race, ethnicity, family income, home language, or disability status—get the education they need and deserve. The goal of NCLB was ambitious, and I readily admit there were challenges in its implementation. That said, I hear from teachers, parents, and school leaders who say they want more ways to understand data from these assessments that will help student learning and provide teachers and parents the information they need to help guide that work and help improve schools.

The accurate, objective, and comparable data collected from statewide assessments shines a light on inequities in our education system and whether growth was consistent for all students. These tests are not perfect, but assessments provide policy leaders and decision-makers with important information on student learning and can be used alongside an array of other measures to effectively allocate additional resources and supports to districts, schools, and students who need them most. Most importantly, assessments provide families with data on their child's learning.

While a well-designed accountability system can highlight educational disparities, a badly designed one can hide achievement and opportunity gaps and enable schools and districts to sweep underperformance—for all students or for individual student groups

—under the rug. In the last twenty years, our nation’s schools have made substantial progress, especially for students from low-income backgrounds, students of color, students with disabilities, and English-language learners. But despite that progress, there is still a long way to go, and the pandemic has further undermined this progress.

At Ed Trust, we believe our students deserve to be at the center of education systems in this country, where their assets are so much more than the system structure, and we can hold systems accountable. The last two years have only deepened this belief. As students lose instructional time to COVID-19 closures, we need to know what states and districts are doing—how they’re spending the billions of new federal dollars—to make a difference in the education of our children, especially Black students like my two boys, Latino and Native students, and students from low-income backgrounds. Using data, we can hold policymakers accountable and improve the education that our students are receiving. Let’s find the best way to do that.

Princess Moss: As we navigate new challenges amidst the COVID-19 crisis, standardized tests are woefully ill-equipped to support the rebuilding of school communities our students deserve. We cannot gain insight into closing opportunity gaps if we continue to rely on myopic, two-dimensional tools that only test a narrow set of skills and subjects. To create safe and just learning environments that will fuel a student’s love of learning, cultivate their independence, and teach them valuable skills, we must employ a diverse array of high-quality assessment methods that truly measure and drive learning.

These challenges didn’t start with COVID-19. I have vivid memories of the years immediately following the passage of No Child Left Behind as an elementary music teacher and local union president in Louisa County, Virginia. Because I taught students of multiple grades, I witnessed how the hyper-focus on test scores in “core” subjects affected students and the educators who taught them. Their joy of learning and teaching faded as curriculum narrowed, and student access to a well-rounded education was sacrificed for increased attention to tested subjects.

As a music educator, I taught to standards and sought focused outcomes that nurtured the development of the whole student. Assessment methods like observation, reflective writing, and performance—tools I used for decades to assess progress—have immense value in supporting authentic learning in ways that high-stakes standardized tests simply cannot accomplish. When I hear from my former students, they do not talk about how the pressure from standardized tests improved their lives. They tell me about how music gave them the confidence to express themselves and how they learned curiosity, patience, and persistence through the songs we practiced together. Our educators want to do more than teach to a standardized test, and they know how if only we let them.

If the use of high-stakes standardized tests improved outcomes for our students and schools, we would have seen the impact by now. After two decades, I hope policymakers realize that standardized test scores do *not* tell us enough about our students’ needs and accomplishments. Ensuring equitable opportunities for all students requires moving away from the belief that we can test students into improvement. It is time to start using meaningful measures of student and school success by working with those who know students best—their educators, families, and communities—

to ensure exceptional schools for all students of every race, place, and background.

Paul Reville: This question about our learnings from high-stakes testing reminds me of a similar question, decades ago, about school-based management. The answer to both was something along the lines of, “Great idea. Love the theory, but we didn’t really ever implement it.” The rarely implemented practice of high-stakes testing, originally proposed as a measurement tool at the heart of an accountability system, was spectacularly uneven and inconsistent with the theory of standards-based reform, which, itself, was never fully integrated at all levels of the education system. Here are several conclusions I draw from the experience:

First, accountability measures like high-stakes testing cannot be solely focused on students without policymakers, school leaders, and teachers being held accountable for first providing students with adequate opportunities to learn. Much of the testing administered in the US proceeded on the assumption that if children were held accountable via stakes attached to some aspect of the testing (real stakes were the exception, not the rule), then adults would modify their strategies and behavior to help students meet the standard. This never came close to happening at scale.

Second, Americans are averse to stakes in the education system, which, prior to standards-based reform, had long been largely unaccountable for performance. When stakes were attached to tests, particularly if those stakes affected adults, there was powerful resistance to any consequences for performance. Standards-based reform was an attempt to impose accountability on a heretofore unaccountable field. Notwithstanding pockets of success, this strategy enjoyed only modest success overall.

Third, adults are much better organized and effective in resisting stakes than children.

Fourth, testing, by itself, is only a measurement tool. You don’t fatten a cow by weighing it. You don’t get better in most organized human endeavors without measuring progress and making it matter.

Fifth, the instrument of testing, embedded in a framework of high standards with a commitment to developing high-quality teaching, is a worthwhile tool in bringing learning and instruction front and center to the debate of what must be done to prepare children for success and better serve those who the education system has most egregiously failed.

And finally, don’t throw the baby out with the bath water. In spite of the failures of some testing regimes, some form of testing and some consequences will be necessary, though alone these are not sufficient tools for achieving excellence and equity in American education.

Diane Ravitch: I gained a new perspective on standardized testing as a result of spending seven years on the National Assessment Governing Board, to which I was appointed by President Clinton. Before my service on NAGB, I naively believed that such tests were fair and objective ways of measuring student performance. Because of what I learned, I no longer believe that. The NAEP tests have one great advantage, from the perspective of teachers and students: they have no consequences. They are “no-stakes” tests because no student takes the entire test, and no school gets a score. The tests are a snapshot of performance that compares all states and many

cities. We can learn from NAEP everything that policymakers want to know: the changes in performance over time for students by race, ethnicity, gender, special education status, and English-language proficiency, as well as achievement gaps among different groups.

NAEP is very different from the standardized tests introduced by No Child Left Behind. The annual tests in reading and mathematics that all children in grades from three to eight must take have no instructional or diagnostic value. The tests are given in the spring, and the results are reported months later, when the students have different teachers. The results are test scores and rankings: they do not tell teachers what students do or do not know. Instructionally, they are useless. Students and teachers learn what percentile the student is in, or whether they tested above or below “advanced” or “proficient” or “basic,” but they do not learn anything about the individual students’ strengths or weaknesses or where they need help. Furthermore, the teachers are not allowed to see the questions or the answers; they never find out how individual students performed on specific questions. They have no diagnostic value. Test publishers treat their questions as proprietary and have vigorously pursued and legally threatened anyone who dares to disclose the content of the tests.

In contrast, if teachers teach a unit in reading or math, they can write tests that diagnose what students learned and what they did not learn. The teacher knows what was taught. The teacher can then give students additional attention in the areas where he or she didn’t understand what was taught.

As a board member of NAGB, I was frustrated to discover that the tests accurately measured the income of students’ families. Students from families with high incomes got the highest scores. Students from families with low incomes got the lowest scores. My curiosity was piqued, so I investigated and learned that every standardized test—be it the SAT, the ACT, or international tests—produces rankings that are closely correlated to family income.

It is important to understand that standardized testing changes nothing, although it does discourage students who get low scores year after year. Since teachers learn nothing about individual students and what they know and can do, the tests provide no guidance about how to help them do better next time.

Holding teachers and schools “accountable” for test scores has been a losing proposition. Given the fact that test scores reflect family income, teachers working in schools with many low-income students will be held accountable for factors beyond their control. Professional societies, like the American Statistical Association, have pointed out that “teachers account for about 1 percent to 14 percent of the variability in test scores, and that the majority of opportunities for quality improvement are found in the system-level conditions. Ranking teachers by their VAM [value-added models] scores can have unintended consequences that reduce quality.”¹

Holding schools “accountable” for test scores has led to the closing of hundreds, perhaps thousands, of schools in low-income neighborhoods. There is no evidence to my knowledge that students got higher test scores because they were moved to other schools.

¹American Statistical Association, “ASA Statement on Using Value-Added Models for Educational Assessment,” April 8, 2014.

Rather than imagine that standardized tests will lead to instructional improvement, why not reduce class sizes so that children who need individualized help can get it? Why not develop community schools where children and families can get free medical care, dental care, and whatever social services they need?

Here is a medical analogy that explains my view of standardized testing. Imagine you have a sharp pain in your abdomen, and you go to a doctor who specializes in gastrointestinal medicine. The doctor gives you a battery of tests. She says at the end of the testing, “Come back to see me in four months, and I will have answers for you.” When you return four months later, the doctor tells you that you are in the 54th percentile of patients who have that kind of pain in their abdomen. She gives you no medications because she learns nothing more from the tests other than your ranking.

NAEP can tell us whatever we need to know about how students in different states and cities are doing in the subjects tested. If we want to know how our own son or daughter is doing, the best way to find out is to ask their teacher. He or she knows your child, knows what your child is capable of, knows if your child is engaged in schoolwork and thriving, knows if they are disengaged, and knows how you can help your child. I trust the judgments of teachers more than big testing conglomerates.

Denise Forte: Yes, we can and must do better to improve the systems that educate our students. But we can’t dismiss what was uncovered through assessments and accountability systems about the generational failings of states and school districts when it came to Black and Latino students’ academic progress.

The first decade after the passage of No Child Left Behind, achievement gaps between Black and White students were closing. Having data and transparency made people pay attention. We learned that schools that excel on state-level assessments are ones with strong leaders and teachers who instruct using an engaging, high-level curriculum that is aligned to college- and career-ready standards. In these environments, teachers didn’t need to “teach to the test,” because high-quality curriculum and assignments are part of the school culture.

The Great Recession changed things. Between 2008 and 2012, the K-12 public education system lost nearly 300,000 jobs, the largest reduction in our nation’s history. Of the jobs lost, over 120,000 belonged to elementary and secondary teachers. These layoffs disproportionately impacted schools serving students of color and students from low-income families. Even when districts were able to reinstate classes and programs, there were still teacher shortages, especially in math, science, and special education. Today, there are even fewer public school teachers than in 2008. Now, with the additional stresses faced over the past two years during the ongoing COVID-19 pandemic, it’s expected that teacher shortages will further increase, and the brunt will be born on Black and Latino students and students from low-income families.

A single assessment will not remove or eliminate the barriers present in the system. We need diagnostic, formative, and summative assessments to understand the full weight of progress. Systems of assessments must be seen as a measure of how well (or not well) state and local systems are providing opportunity for all

students, including students from different racial and low-income backgrounds, and utilized to shine a light on and direct resources to the places that need additional support. I would argue that summative assessments, when built and delivered well, never lead to “drill and kill” pedagogy or change the focus of high-quality instruction. However, I recognize that this has been a real consequence for schools that have not had the resources or capacity to provide all students with high-quality core instruction. That’s why it is so important that assessment data is publicly available to help target federal, state, and local resources toward improving district and school capacity.

Some have said we should rely more on grades than summative assessments. As a mom of two boys in elementary and middle school, I agree grades can tell parents a lot about the day-to-day engagement and performance of their individual students. However, grades from one school cannot be compared to grades from classroom to classroom, school to school, or district to district, nor can policymakers fully depend upon a teacher’s evaluation of student work for purposes of accountability and resource allocation. Decades of research have shown that teachers—despite their best efforts—often give racially biased evaluations of student work and that biased evaluations can affect students’ future learning and course-taking decisions. For example, a 2018 study by Nicholas Papageorge, Seth Gershenson, and Kyung Min Kang found that White teachers, who comprise about 80 percent of American educators, have far lower expectations for Black students than they do for similarly situated White students. More recently, a 2020 study by David Quinn found that in a randomized control trial, teachers rated a student’s writing sample lower when it was randomly signaled to have a Black author versus a White author.²

Assessments, along with accountability systems, have a place in our school systems. And the data we get from them can identify the real inequities in our schools within and across districts. State accountability systems can and should serve as a signal of school performance and influence policy and spending decisions.

Princess Moss: High-stakes testing should not dictate high-stakes decisions. The need to move away from our overreliance on standardized tests is not based solely on the fact they have not worked to close opportunity gaps in the past two decades; it’s also because they were not built to support the future our children deserve. Sound assessment practices will help us identify a student’s strengths and areas for growth, encourage students’ love of learning, measure a program’s effectiveness, determine instructional strategies, and inform the creation of appropriate, high-quality learning experiences. Assessment systems that are asset-based, multidimensional, and well-rounded are key to securing a future where each child learns in a caring, inclusive environment that has high expectations for every student.

Assessment that more accurately reflects a broad range of student learning promotes critical thinking and deep subject-matter knowledge and encourages students

²Nicholas W. Papageorge, Seth Gershenson, and Kyung Min Kang, “Teacher Expectations Matter,” Working Paper 25255, National Bureau of Economic Research, Nov. 2018, doi.org/10.3386/w25255; David Quinn, “Experimental Evidence on Teachers’ Racial Bias in Student Evaluation: The Role of Grading Scales,” *Educational Evaluation and Policy Analysis* 42, no. 3 (June 2020), 375-92.

to thrive in their classrooms, communities, and beyond. For example, the New York Performance Standards Consortium—which incorporates teacher-directed, performance-based assessment—has supported student success, particularly among vulnerable student populations, with students demonstrating lower dropout rates and higher rates of college enrollment. The Consortium assesses student work and progress using an array of methods from critical writing to presentations, guided by educators who participate in teacher-led professional development, mentoring, and observation to continuously improve their practice.

Students have much more to show us than just filling in bubble tests. We need to trust our educators and partner with them to ensure their expertise, knowledge, and experiences are inherent in the creation of classroom, local, and statewide assessment. After twenty years of testing to identify opportunity gaps (surprise, they are still there!), decision-makers must recognize that closing our eyes to everything except test scores is not a racially and socially just method of determining student progress or identifying student needs. We must offer our students equitable opportunities to demonstrate their knowledge and skills and provide a robust system of school quality indicators that contextualize data and ensure students have the resources and support they need to thrive.

I think we can all agree that we want what is best for our students and schools. That is why I am so proud to help lead the NEA's Task Force on the Future of Assessment. The Task Force includes educators from across the country who are passionate about transforming assessment to reflect our equitable, robust, asset-based vision. After extensive listening sessions with our members and meetings with psychometricians, researchers, and experts who have helped guide us, the Task Force drafted NEA's Principles for the Future of Assessment. The Principles include collaborating with the community, championing the expertise of educators, prioritizing student self-efficacy, generating and employing well-rounded evidence, and ensuring all students opportunities to participate in culturally relevant and responsive assessment. This is the blueprint we need to ensure that our conversation twenty years from now is a celebration of our progress and not a painful reminder of a decades-long race to the bottom that drained the joy from teaching and learning and placed our education system and our children last in line on the world stage.

Paul Reville: It is time to move on after thirty years of arguing about testing. This conversation has become a distraction from what really matters: finding and embracing strategies to improve children's well-being, learning, and success. Despite all our investments of resources, energy, and good intentions, we have hardly moved the needle in preparing larger proportions of our children to thrive. We're going to have to re-envision both the problem and potential solutions. Schools alone, as currently constituted, are not designed or operating in ways that will ever conceivably approach our ideals of leaving no child behind or ensuring every child's success. Our reform efforts have exposed the fundamental inadequacy of an education system that involves only 20 percent of a child's waking hours and tends to operate in a one-size-fits all, factory modality. Our legacy system is simply too weak an intervention to achieve our ambitious goals for equity and excellence.

As to testing, the argument is stale. The stakes are largely gone, the federal government is steadily relaxing requirements, teachers and many in the general public are fed up with too much testing. It's time to re-balance and re-envision. Less testing, but still measure progress; modest and appropriate stakes; but most importantly, let's broaden our conception of what needs to be measured. Let's revisit "opportunity to learn" and start rating policymakers, education leaders, and communities on what they have or have not done to put in place genuine opportunities for children to thrive and learn. Let's take a more holistic view of what children need to thrive and whether or not they have the necessary supports and opportunities to be successful inside of school and out. Let's look at how we're addressing the myriad factors related to race and poverty that get in the way of children coming to school and being ready to learn when they do. Why not measure the degree to which our communities are providing the opportunities and supports for young people, what they need to thrive outside of school? Let's strike a new social compact and hold all parties accountable.

Instead of dwelling on what we don't like and what doesn't work, let's try to rebuild a consensus on what it'll take to realize our ambitious equity and excellence goals. We have lost what was once a widespread consensus on what it would take to improve education. Former allies like many in the business and philanthropic communities have abandoned the field due to frustration, disillusionment, and pessimism about reform, to say nothing of battle fatigue from all of our internecine battles over subjects like testing, charter schools, and now critical race theory. Alarming, teachers, too, are abandoning the field in droves, leaving a major talent crisis in their wake. Powerful unions have hunkering down to protect their members in the storm rather than leading a movement to re-envision the field. The field of education is a field at risk.

I don't have a formula for a consensus, but I'd love to see the focus of our education discourse shift to constructive, new strategies that might work for children. Here are some promising avenues, many of them having opened during the pandemic, which are worth exploring:

Personalization: Every child should have a success plan and a navigator to help them and their family successfully navigate challenges and have access to supports and opportunities inside and outside of school.

Family Engagement: It's time to lean into the field's rhetoric about families being partners and make those partnerships real and enduring.

Constructive Use of Technology: COVID forced us toward a widespread embrace of tools our field had been reluctant to use. Let's not drop these tools and their affordances as we return to in-person schooling. Make the most of their capacity to enhance learning in so many ways.

Restructure Schools to Build Relationships and Ensure Mental Health: Student mental health was at crisis levels before the pandemic; now matters are significantly worse. Our secondary schools, in particular, are not built to foster relationships among students and between students and faculty. Better relationships are an essential component of any path forward in education.

Deeper Learning: Project-based learning, applied learning, school-to-career pathways, early college, hands-on, experiential learning, problem-solving, collaboration,

dual enrollment—these are all underutilized strategized that work well with students and need to be much more widely available.

Re-imagine the Calendar and Schedule: Differentiate the schedule to meet each child where they are and give them what they need. Break the back of the factory model and shape time and opportunity to provide students what they need to achieve mastery. Make sure all children have access to after-school and summer learning opportunities.

Bear Down on Early Childhood: Make high-quality, early education an entitlement.

Finally, we must think more holistically about what children need to be successful. We need to think like parents and nurture our children as society's greatest asset. We need a default system that enables us to do for all children what those of us who have privilege are routinely able to do for our children in terms of education, support, and opportunities outside of school. It's just not rocket science!

Diane Ravitch: For many years, I believed strongly in the value of standardized testing and accountability. However, the more I learned about the tests and about their negative effects on students, the less I trusted them. The tests used today to rank students or evaluate their teachers are neither reliable nor valid. Tests accurately measure the students' socioeconomic background. Children of privilege consistently get high scores, while children of low-income families consistently get low scores. Perhaps we should give more attention to the causes of academic performance than to test scores.

We should have learned by now that children will do better in school if they are healthy, well nourished, and live in homes that have the means to meet their needs. The cause of test score gaps is not teachers, teaching methods, or curriculum, but the well-being of students. Achievement gaps reflect economic gaps. As a society, we have naively believed that raising test scores would reduce poverty, and we attack teachers and schools when our fantasy does not happen.

Standardized tests are normed on a bell curve. By definition, a bell curve has a bottom half and a top half. The bell curve never closes. Children of privilege dominate the top half, children who live in poverty dominate the bottom half. A few will rise or fall irrespective of their circumstances, but they are outliers.

Standardized tests do not improve academic skills, nor do they provide useful information to parents or teachers. Under the current regime of standardized testing, neither teachers nor parents learn anything about students' individual performance. The students get a ranking (advanced, proficient, basic, below basic, or some variation thereof), but the teacher never learns how individual students answered specific questions. The tests have no diagnostic value. Teachers never learn which questions the students answered correctly and which they answered incorrectly.

Teachers want to know what their students understood and what they missed. The tests don't tell them, and students and teachers are forbidden to reveal test questions or answers.

If the tests have no diagnostic value, of what value are they? They give us the illusion that we are "doing something," when in fact we are using an instrument that is so important—even though ultimately of no value—that it crowds out the arts, play, history, civics, and everything else that is not tested.

We have wasted billions on testing over the past two decades that might have been spent to reduce class sizes, provide free medical care, and supply three nutritious meals a day to needy students.

It is past time to embrace a new vision for helping students lead better lives.

Denise Forte: I want to echo the sentiment that it's time to rebuild a consensus to reach our common goals for supporting students who have long been furthest from opportunity. Others are attacking the very core of our educational system. Our schools have become the battleground for the soul of our nation. It's not just that right-wing ideologues are attacking students' right to learn the truth about America, using the pretext that our nation's educators are teaching so-called critical race theory. It's not just that state legislatures began banning what they called "divisive concepts" when it comes to teaching about race, or that books are being pulled from school libraries and reading lists across the country. Or that social and emotional supports are under attack because they could be "indoctrinating" students.

Let's call this all what it is: an attack on public education and an attack on students of color. And the result will be a further disparity in the resources available to students, particularly those of color or living in poverty. Now is the time to fight even harder to make sure that students—particularly for those whom public education has failed for far too long—receive a high-quality education that prepares them to live a life of their choosing.

And, we should start with this shared premise: All children deserve a public education system that offers high standards for all, and that system should be able to tell us whether or not it is meeting this goal. Statewide summative assessments are the common yardsticks for measuring progress and help expose gaps that are otherwise covered by looking at averages. Other assessments, formative and diagnostic, should also be used to offer a comprehensive view of student performance. Standards are shared expectations for what all students should know and be able to do. Together, standards and assessments tell us a story and help us understand where students are at a point in time and can be used to direct resources into communities with persistent gaps.

Without a comprehensive system of standards and assessments that offers a guidepost for teaching and learning, it becomes even easier for opponents of an equitable education system to mount and succeed with attacks on public education. If there is no expectation for all students in a state to achieve at a high standard and we lack data on whether this is happening in schools and districts across the country, then misguided leaders can continue to distract by banning books and creating "tip lines" to report on teachers. Without data on whether a student is reading on grade level or has met the competencies for eighth-grade algebra, it will be even easier for policy leaders to dismiss the real challenges facing Black and Latino students.

I would agree with Diane that NAEP is an important tool. Because it uses sampling and is voluntary for states and students, it lacks what is currently required in a statewide assessment system and does not offer enough information for district or state leaders to drive system improvements. And herein lies the ongoing challenge. If we are to address the systemic equity issues in education that are present across the country, then we need a comprehensive system for doing so. As I wrote earlier, while

we all can agree that the current assessment system is not perfect, it does remain the best tool we have right now. And, we should all be working together to create the model future assessment.

What should that look like? I look forward to working alongside the other authors/commentors to build and grow a public education system where the focus moves away from “fixing kids” and toward high expectations for all students. We must center school and district policies that create equitable learning environments as well as ensuring that students have access to stable housing, nutritious meals, and the many other supports they need to succeed in the nation’s classrooms. And, if this is done well, the data we obtain from a high-quality and comprehensive system of assessments will drive school improvement and better allocate and use resources—people, time, and money—to create student experiences that enable all children to achieve, particularly those of color or living in poverty.

Princess Moss: I think we all agree that it’s time for us to work together to create a common and constructive vision that will ensure high-quality opportunities for every student. Let me be clear: the NEA supports student-centered assessment methods that measure learning comprehensively, and not just by filling out bubbles. We need assessment that helps inform instruction and doesn’t take away from it. There are already examples of assessment that work, and though not perfect, they provide actual, usable results for students and other stakeholders. For example, we support the administrations of the National Assessment of Educational Progress (NAEP) and the Programme for International Student Assessment (PISA), which, over time, provide data that can give us important information about the efficacy of our policies. Simply put, we believe that we can improve assessment policy and practice to better support teaching and learning. And we believe that improvements should be made in partnership with the educators who are closest to the students and know them best.

Like everyone on this panel, the NEA is not new to this work. Over the last twenty years, we have joined our voices with hundreds of renowned organizations to call for improvements to assessment and accountability policy. We assembled a Task Force on the Future of Assessment in 2021. And over the last year, we rallied current and potential partners around a common vision and created a set of Principles for the Future of Assessment. Our hope is that this document reflects the voices of all who have been involved in this process, and I invite my fellow panelists to review what we have developed thus far, because I believe it provides further evidence that we are all working toward a common goal. For example, Paul mentioned that we should dramatically increase students’ participation in more authentic assessment methods. Likewise, Diane noted that teachers can provide insightful, detailed information to parents and families about students’ academic successes and areas for growth. And Denise painted a vivid picture of some of the red-herring headlines in the news today that distract from bona fide efforts to close deeply exacerbated opportunity gaps as we emerge from the pandemic.

We believe in the transparency and answerability of our education system to the students and communities it serves. As practitioners, our positions are informed by living out the real-world ramifications of policy every single day in the classroom. We’ve seen firsthand the shift toward “teaching to the test” that is both a distraction

from and disservice to our students' learning experiences. NEA's job as a professional association is to advocate for a seat at the table for our members in the profession they love. The nearly three million educators that make up our membership are so much more than test proctors—they are dedicated professionals who truly understand our nation's students and their growth. We are champions of public education who—despite unprecedented challenges, workloads, and stress—are passionate about our jobs and ensuring that each one of our students has access to a quality public education that develops their potential, self-determination, and character.

To center our assessment systems on the needs of our students, we must encourage our educators to use their professional knowledge and experience, autonomy, and skills to employ a variety of measures to accurately assess student growth. The NEA is fully committed to supporting our nation's educators and leading the charge. Together, we can create, implement, and evaluate quality assessment processes that advance inclusion and clearly communicate actionable results to our students, families, and communities. The Principles for the Future of Assessment outlines this new paradigm, and we welcome our fellow leaders in education policy to provide feedback and join us as we work to realize this vision.

Paul Reville: From the distinct voices in this dialogue, I'm encouraged that some common themes emerge: a commitment to equity, the need to broaden and deepen our approaches to measuring student learning, the need for improving both diagnostic and accountability features of our current testing regimes, faith in teacher judgment coupled with the belief that cross-cutting and criterion-referenced measures enable the identification of pockets of inequity, caution about stakes, a principle that various adults in our policy and education systems need to be accountable for putting in place "opportunities to learn" before student achievement is measured, a basic understanding that you can't achieve goals without measuring progress, and general agreement that formative and summative evaluation is fundamental to the education process. And, of course, there's the underlying conviction, seemingly shared by all of the writers, that American public education is failing to achieve its twin ideals of excellence and equity, thus leaving far too many children well behind the standard needed for them to achieve success in this society. That is the tragedy, especially in a nation of great wealth.

As I suggested in my last response, I think we need to look beyond testing, and beyond education, to diagnose what is undermining the American Dream of equal opportunity. It isn't just failing schools. That's just one symptom of a failing society and an economy that so unfairly distributes wealth as to undermine and marginalize large segments of the population who are forced to live in conditions that barely support well-being or learning. This is what keeps our "nation at risk" much more so now than when the nationally influential report of that title, *A Nation at Risk*, was issued in 1983.

In our field of education, we have often, and for a long time, been prime exemplars of the adversarial, divisive, invective-filled, hostile discourse that is now dividing our entire society. Our destructive, internal wars over testing, charter schools, standards, phonics versus whole language, inclusion, bilingual education, etcetera have set the

stage for current death spiral debates over masking, vaccines, and critical race theory. Now, the public education sector itself is at risk.

It's become a cliché to rationalize any position on any education issue as “what’s best for the kids,” but I keep hoping that we as educators can come together, just this once, to make common cause with community, children’s rights, and civil rights leaders, to advocate for the essential supports and opportunities that all children need to be healthy and ready to learn. We should be “all in” and directing our collective muscle to support the extension of the child tax credit, health and mental health care for all, the elimination of hunger, plus universal access to early childhood care and education, to name just a few policy and budgetary changes that need to be made if America has any hope of restoring social mobility and becoming an equal opportunity society.

I’m afraid that if we get too consumed by our own internal battles, we’ll fail to see, let alone meet, the larger challenges that face our children, their families, and our society. This nation’s failure to fairly distribute prosperity and well-being has given us the divided nation we see today. Tomorrow, that failure could very well lead to the demise of our democracy. We, as educators, can make a difference.

Additional Readings

- Boudreau, Emily. “Leaving No Child Behind: How Education Redesign Lab’s Student Success Plans Helped Districts Meet Needs during the Pandemic.” *Harvard Graduate School of Education*, Oct. 1, 2021, <https://www.gse.harvard.edu/news/21/10/leaving-no-child-behind>.
- “Deep Dive: Lifelong Learning: Cradle to Career.” *Thriving Together*, June 2020, <https://thriving.us/deep-dive-cradle-to-career/>.
- The Education Trust. “Students Can’t Wait: Resources.” Mar. 25, 2022, <https://studentscantwait.edtrust.org/>.
- The Education Trust. “Students Can’t Wait: Resources.” “New School Accountability Systems in the States: Both Opportunities and Peril.” Mar. 25, 2022, <https://edtrust.org/wp-content/uploads/2014/09/AccountabilityOverview.pdf>.
- Harvard Graduate School of Education. “EdRedesign Receives \$3.2 million in Inaugural Grants to Launch Institute for Success Planning” (press release). Mar. 4, 2022, <https://edredesign.org/files/edredesign/files/sp-institute-pr-pdf-final.pdf?m=1646855726>.
- Jokic, Marina. “EdRedesign to Launch Institute for Success Planning.” *Harvard Graduate School of Education*, Mar. 4, 2022, <https://www.gse.harvard.edu/news/22/03/edredesign-launch-institute-success-planning>.
- National Education Association. “2020 NEA Policy Playbook for Congress and the Biden-Harris Administration.” Nov. 12, 2020, <https://www.nea.org/resource-library/2020-nea-policy-playbook-congress-and-biden-harris-administration>.
- National Education Association. “Learning beyond COVID-19: A Vision for Thriving in Public Education.” Mar. 5, 2021, <https://www.nea.org/resource-library/learning-beyond-covid-19-vision-thriving-public-education>.
- National Education Association. “National Education Association Policy Statements 2021-2022.” Mar. 25, 2022, https://www.nea.org/sites/default/files/2021-09/NEA%20Policy%20Statements%202021-2022_0.pdf.
- National Education Association. “NEA’s Principles for the Future of Assessment.” Feb. 4, 2022, https://neahq-my.sharepoint.com/:b/g/personal/cdonfrancesco_nea_org/ETLtyTlkGrVNoquy_Np8aK0Bol1Rof6sTivA14BvCpkbxQ.
- Olson, Lynn, and Thomas Toch. “Changing the Narrative: The Push for New Equity Measures in Education.” *FutureEd*, Nov. 2021, https://www.future-ed.org/wp-content/uploads/2021/11/FUTUREED_REPORT_EQUITY.pdf.
- Reville, Paul. “American Schools Need a New Paradigm: Personalization.” *Boston Globe*, Sept. 13, 2021, <https://www.bostonglobe.com/2021/09/13/opinion/americas-schools-need-new-paradigm-personalization/>.

- Reville, Paul, and Lynne Sacks. "A Strategy for Re-engaging Students Post-Pandemic." *FutureEd*, Mar. 1, 2021, <https://www.future-ed.org/a-strategy-for-re-engaging-students-post-pandemic/>.
- Reville, Paul, and Geoffrey Canada. "The Time Has Come for Truly Personalized Learning—With a Navigator to Make Sure Each Child Succeeds." *T74*, April 12, 2022, <https://www.the74million.org/article/reville-canada-the-time-has-come-for-truly-personalized-learning-with-a-navigator-to-make-sure-each-child-succeeds/>.
- Spurrier, Alex, et al. "The Impact of Standards-Based Accountability." *Bellwether Education Partners*, Summer 2020, https://bellwethereducation.org/sites/default/files/Bellwether_Accountability-Impact_Final.pdf.
- Ushomirsky, Natasha, David Williams, and Daria Hall. "Making Sure All Children Matter: Getting School Accountability Signals Right." *The Education Trust*, Oct. 2014, https://edtrust.org/wp-content/uploads/2013/10/All_Children_Matter.pdf.