



NOVEMBER 2021



ASSESSMENT BOOTCAMPS



KEY FINDINGS



CHARLES BARONE, Vice President of K-12 Policy,
Education Reform Now

NICHOLAS MUNYAN-PENNEY, Senior Policy Analyst
Education Reform Now

ASSESSMENT BOOTCAMP

1

Assessment Types, Their Uses, and Building Balanced Assessment Systems

FEATURING:



Charles Barone

Vice President of K-12 Policy
Education Reform Now



Susie Feliz

Vice President of Policy and Legislative Affairs
National Urban League



Stuart Kahl

President
Kahl Balanced Assessment Practices



Paige Kowalski

Executive Vice President
Data Quality Campaign

ASSESSMENT BOOTCAMP

2

Innovations in Statewide Assessments and Accountability

FEATURING:



Laura Jimenez

Director, Standards and Accountability
Center for American Progress



Phillip Lovell

Associate Executive Director
Alliance for Excellent Education



Chris Minnich

CEO
NWEA



Jennifer Randall

Associate Professor Director of Evaluation for the Center for
Education Assessment
University of Massachusetts, Amherst

ASSESSMENT BOOTCAMP

3

Making the Case for Statewide Summative Assessments

FEATURING:



Roxanne Garza

Senior Education Policy Advisor
UnidosUS



Lindsay Kubatzky

Policy Manager
National Center for Learning Disabilities



Tanji Reed Marshall, Ph.D

Director of P-12 Practice
The Education Trust



Keri Rodrigues

Founding President
National Parents Union

Summative, statewide assessments have a unique and irreplaceable role, particularly with regard to being an global indicator of system quality and providing achievement data that is aligned with state standards and that can be compared across districts and different groups of students.

- **This first session** was designed to provide advocates and policymakers with a shared language and knowledge base about different types of assessments and the purposes they are—and are not—designed for and how these various assessments can work together to create coherent, balanced assessment systems.
- **The second session** was devoted to discussing potential changes, improvements, and innovations in state testing systems, including the creation of anti-racist assessments.
- **The third session** was dedicated to making the case for statewide assessments as a tool for providing information on the performance of state, district, and school-level education systems, for identifying and addressing achievement and opportunity gaps, and for channeling resources and support to schools in need of improvement.

Several key themes emerged from these discussions:

- **It's important to distinguish between different types of assessments** and to be clear as to the purposes each is designed to serve.
- **Statewide, summative assessments have a unique and irreplaceable role**, particularly with regard to being an global indicator of system quality and providing achievement data that is aligned with state standards and that can be compared across districts and different groups of students.
- **There is near consensus that our current assessment system has inherent biases when it comes to race**, but there is a divergence of opinions on the extent of the problem and what should be done to resolve it.
- **Confusion about the role that assessments play in accountability systems** may be driving much of the opposition to statewide summative assessments.
- **Within each of these areas, there are a variety of different innovations in assessment systems that are being piloted throughout the country**. These vary from relatively minor adjustments to more dramatic overhauls, each with their own advantages, disadvantages, and tradeoffs.



Distinguishing Between Different Types of Assessments

In the first session **Paige Kowalski, Executive Vice President at Data Quality Campaign**, pointed out it's a lack of understanding about the discrete purposes of different assessments that can sow confusion—and produce data that doesn't answer the questions we want answered. "Different assessments yield different types of results and for different purposes and the trick is understanding what you're going for—what kinds of questions you're trying to answer—so you can pick the right assessment instrument."

One useful taxonomy distinguishes between summative assessments, formative assessments, and diagnostic assessments:

- Summative assessments provide an end-of-year snapshot of how students are performing against grade-level standards, providing data that is comparable across student groups, schools, and districts within a state to aid policymakers and schools and district leaders in making decisions about resource allocation and program improvement.
- Formative assessments are conducted periodically in the classroom by educators to provide critical data about student understanding of current content, allowing educators to make real-time adjustments to instruction, grouping, and instructional supports.
- Diagnostic assessments are typically external assessments and provide educators and district leaders with information about student achievement relative to both grade-level and non-grade-level content, which allow schools to make student class and/or group placement decisions and teachers to plan instruction.

*"Different assessments yield different types of results and for different purposes and the trick is understanding what you're going for—what kinds of questions you're trying to answer—so you can pick the right assessment instrument."
—Paige Kowalski,
Data Quality Campaign*

Statewide summative assessments are essential for answering questions for program evaluation and school improvement. **Stuart Kahl, President of Kahl Balanced Assessment Practices**, noted that, "You have an awful lot of comparisons that are allowed by state tests, and the results should raise questions like: Why are we not performing as well as other schools in our state that serve the similar population of students? Why is this subgroup of ours not performing as well as the same subgroup at other schools across the state?"

Phillip Lovell, Associate Executive Director at All4Ed, started the second bootcamp session with an illuminating metaphor for the unique role of statewide summative assessments. They "provide a check on the system," he said. "It's similar to a 'check engine' light on the dashboard of your car. When the check engine light comes on it

doesn't tell you exactly what's wrong or how to fix it. But it does tell you there's an issue....It doesn't do more than that, but it's an important function."

Susie Feliz, Vice President of Policy and Legislative Affairs at National Urban League, added that statewide summative assessments are also about "ensuring that all students are taught to the same high standards and the statewide assessments are aligned to the states' college and career readiness standards," while local assessments—though valuable for addressing other needs—are not necessarily aligned to these standards.

Lindsay Kubatzky, Director of Policy and Advocacy at National Center for Learning Disabilities, similarly cautioned against using statewide summative tests for the more fine-grained work of diagnosing specific student learning needs or directly informing day-to-day adjustments to instruction. "There shouldn't be any decision making happening from [statewide summative] assessments at the student or educator level. This is really about school systems in the aggregate and it's really important that states and districts use these summative assessments to target resources."

Yet **Chris Minnich, CEO of NWEA**, noted many states and districts are, in fact, using statewide summative assessments for the purposes Kahl warns against. Since testing vendors like NWEA and others are ultimately responsive to their clients—states and districts—Minnich said summative assessments have had to expand in scope and depth to accommodate somewhat conflicting purposes.

"We're asking statewide assessments to do a lot more than they were intended to do," Minnich stated. "The biggest problem that we have to resolve is purpose. If we decide as a country that we want it [summative statewide assessments] to be that "check-engine light," then we would design a very different assessment."

"There shouldn't be any decision making happening from [statewide summative] assessments at the student or educator level. This is really about school systems in the aggregate and it's really important that states and districts use these summative assessments to target resources."
—**Lindsay Kubatzky**
NCLD

Innovation Focused on Assessment Purpose

Laura Jimenez, Director of Standards and Accountability at Center for American Progress, agreed, saying that determining a clear purpose for summative assessments is a critical first step as more states move to develop new, innovative assessments. "There are two overarching purposes that annual, evaluative assessments can take. There's a check on the system and then there's a purpose more connected to driving teaching and learning."

Jimenez, referring to content from her recent [series on innovations in assessment](#), said that if states opt to have solely a systems check function, statewide summative assessments could move to limit administration time, and thus decrease interruptions to instruction, while still allowing for comparability, by moving to a matrix sampling of items. This would entail each student taking a

representative sample of items from a larger assessment, meaning you could say “all students took this entire assessment collectively, but no single student took the entire assessment.”

Conversely, Jimenez noted that if states wanted to move to a model more focused on informing instruction, they could follow the lead of what a handful of states ([Georgia](#), [Louisiana](#), and [North Carolina](#)) are doing under the federal Innovative Assessment Demonstration Authority (IADA) pilot program by developing through-year assessments. In these assessments, students are given shorter assessments a few times throughout the year, results are returned to educators quickly to help adjust instruction, and summative items embedded in each assessment are aggregated to give each student a final score at the end of the school year. Also part of IADA, [New Hampshire](#) is developing more in-depth performance assessments that are directly connected to what students are learning in the classroom.

Finally, Lovell, reminded viewers that as states and testing vendors work to innovate they need to not lose sight of why they are innovating and that summative assessments are just one part of a balanced assessment system. “We have to be mindful of what problem we are trying to solve. Is it limiting test time? Is it using the tests more for instruction? ...There are different assessments for different purposes, no one assessment should rule them all.”



Throughout our conversations, our expert panelists highlighted that it's the ability to make apples-to-apples comparisons which make statewide summative assessments so valuable.

Kahl noted, "Test results are only meaningful in comparison to something: in comparison to a pre-established standard, in comparison to previous performance, or in comparison to some other groups...They give you an external perspective on how well your program is doing."

Similarly, Kowalski emphasized that "statewide summative assessment data is the only data point that a state has to look out across their own districts to do program evaluation, to understand how groups of students are doing, to understand which high schools are preparing their students for college or careers."

"Statewide summative assessment data is the only data point that a state has to look out across their own districts to do program evaluation, to understand how groups of students are doing, to understand which high schools are preparing their students for college or careers."

*—Paige Kowalski,
Data Quality Campaign*

And Feliz very clearly said the importance of statewide summative assessments in advancing educational equity, "It's really about shining the light on academic disparities, particularly for students of color and other vulnerable students. ...For Urban League affiliates and their civil rights partners on the ground, the use of standardized tests has helped reveal these long standing racial disparities in academic opportunity which has armed them with the evidence that they need to advocate for change."

Without statewide summative assessments, these disparities wouldn't be visible. The resulting growth data is, Kowalski notes, "our only comparable data point about performance and it is our only equity indicator on performance in existence right now."

Tanji Reed Marshall, Director of P-12 Practice at The Education Trust, noted how these comparative data are also critical for local administrators, "Building leaders and district leaders really need to understand, from a comparative standpoint, where students are in their districts. ...If there are districts where schools have similar demographics and one school is performing well with a group of students then the question becomes if there is a school with similar demographics where students are underperforming, how then can we make the changes necessary?"

Why Local Assessments Can't Replace Statewide Tests

Asked about states who have expressed interest in using local assessment data in lieu of statewide summative assessments, Kahl stated that "In my mind, whether it's for accountability or for program evaluation and improvement, that lack of comparability is a serious problem with that approach." Kowalski added that "You could have a state with 500 districts and 500 district scores sent up to the state and they can publish that but what is that data set? What can you do

with it? What questions can you really answer? And it will get misused because it's not even apples and oranges, it's apples and eggplants."

The panel's moderator, **ERN's VP of K12 Policy, Charles Barone**, pointed out that without statewide summative assessments, "you can have some districts look good because they set a lower bar and other districts look not as good because they set a higher bar... Without the apples-to-apples comparability you really can't do the resource part of this in any effective way. You'd actually have a fairly regressive distribution of resources."

Roxanne Garza, Senior Policy Advisor for UnidosUS, agreed, also noting that "Local, formative, interim, or diagnostic assessments might not be aligned to grade level expectations. They might not provide the same level of support or accommodations for students with disabilities or English learners. And they likely cannot help state leaders identify systemic inequities across districts... We understand that assessments are not perfect. They alone are not enough to address inequity, but they must be paired with real action so that those inequities are addressed. So we do see the statewide, summative assessment as an important first step."

"We understand that assessments are not perfect. They alone are not enough to address inequity, but they must be paired with real action so that those inequities are addressed."

*—Roxanne Garza
UnidosUS*

Education Stands Alone in the Push for Less Data

Keri Rodrigues, Founding President of National Parents Union, stated the importance of summative assessment data in stark terms with a comparison to the criminal justice system. "When we take a look at FBI data and the things that they test for, it's murder rates, it's assaults, break-ins, arson, stop and frisks, arrests, it's not everything that's in the criminal justice system but it's the data that we have."

"Now we would never, especially where we are in our American society and conversation right now, ever have a conversation about saying 'this data makes us feel really uncomfortable and upset, so you know what we're not going to look at this data anymore or instead we're going to allow the cops to assess themselves and run their own evaluation process but we don't them to feel bad about what the data might tell them.'"

"Yet in the education system," Rodrigues said, "which we have all agreed is mired in generations of institutional racism, and then we're having a conversation about whether or not we need data, whether or not we should assess and have information to make decisions on and determine inequities? It's crazy that we're even spending time debating that. Of course we need data."

"In the education system, which we have all agreed is mired in generations of institutional racism, and then we're having a conversation about whether or not we need data, whether or not we should assess and have information to make decisions on and determine inequities? It's crazy that we're even spending time debating that. Of course we need data."

*—Keri Rodrigues
National Parents Union*



Racial Bias in Assessments

While many view assessments as a key tool for advancing racial equity in our schools, there was ample discussion across all three webinars about the extent to which assessments are racist or racially-biased and what should be done to address that.

Marshall did not mince words here. “Are tests inherently racist? Yes. Full stop.” She notes, however, that doesn’t mean educators shouldn’t prepare students of color for these assessments. “The inherency of racist principles underlying an assessment does not negate the fact that a child must pass it anyway.”

Randall similarly noted that the critical context here is that “the field of educational measurement itself was founded within the eugenics movement,” but goes further to say that “this history of white supremacy, racism, oppression continues to haunt the field and our practices, our policies, our guidelines, and our assessments even today.”

She provided key examples of racial violence within assessments and their development, as she outlines in her [recent work](#):

- First, “When we have the assumption, for example, that we must assess all students in so-called standard English, and standard English alone, because it is formal English and other linguistic formations such as African-American English is informal and inappropriate, that’s racial linguistic violence.”
- Second, “When we have bias, fairness, and sensitivity guidelines for large scale assessments like Smarter Balanced that explicitly forbid the inclusion of content related to racial injustice because it doesn’t want to cause bad feelings in students that’s white supremacist experiential violence.”
- And finally, “[assessment developers] are basically required to ignore the fact that Thomas Jefferson owned slaves and he was a rapist. Instead, all we post on assessments is information about how great Thomas Jefferson was and that he wrote the Declaration of Independence. This is an example of racist, whitewashing historical violence.”

Kahl, who has been involved in the design of assessments for decades, pushed back on the idea that statewide assessments are biased by design. “From the beginning we’ve had multiple steps in the process of test development to ensure [cultural bias] doesn’t happen. We have bias and sensitivity reviews. But even before we get to that stage of committees looking for just those things that can lead to issues and differential performance of kids, the test developers are trained on all these factors they have to avoid in these items, so there a great deal of effort in the development process to make sure that tests are unbiased.”

Kahl also pointed out that “after field testing of items (before final tests are produced), testing companies perform statistical analyses to identify items that are biased against particular groups. Such items are not used in the final tests. These analyses identify items that show differential performance between student subgroups that are equal in or statistically equated on ability. That would be bias. Thus, the final operational tests

detect the extent of real differences in performance—they yield data that are necessary to monitor achievement gaps and the effectiveness of efforts to reduce them.”

Yet, Rodrigues stated that these processes may not always be effective, “Here in Massachusetts we have two panels of educators that are supposed to be vetting these tests for diversity, equity, and inclusion. And still every year we have these arguments, things get through that process.”

Systemic Racism Extends Beyond Assessments

Garza, meanwhile, argued against what she sees as the flawed logic of assessment opponents, while agreeing with their point about bias within assessments, because assessments are “just one piece of a broader education system that is systematically racist. ...We can definitely continue to improve assessments and make them less biased, but we can’t really expect the school system to improve outcomes for kids just by getting rid of statewide summative assessments. That’s not gonna happen. If anything the assessments test the inequities in the system.”

Rodrigues agreed, noting “We have to acknowledge that 80% of the teachers leading our classrooms are white women. They have a particular lens. We have not done the work of addressing implicit bias that exists within our classrooms and within our systems.”

And she said, “We can’t just assume positive intent because even when people are approaching racism, they’re doing it through a positive ‘Well, these poor babies, they’re too poor, so I don’t want to stress them out.’ Let me tell you about what’s stressful. When you graduate from high school and are given a diploma and then you try to get into college—even community college—and you have to take two years of remedial courses before you take a college-level class because what you were taught wasn’t on grade level and you do not have proficiency.”

“We can definitely continue to improve assessments and make them less biased but we can’t really expect the school system to improve outcomes for kids just by getting rid of statewide summative assessments. That’s not gonna happen.”
—Roxanne Garza
UnidosUS

Addressing Bias in Assessment

When it comes to addressing bias in assessments, Minnich pointed to the assessments NWEA is currently piloting in Louisiana where students are being assessed on the same pieces of literature they have already learned about in the classroom. These assessments are designed to level the playing field on student content knowledge which has been shown to correlate closely with scores on assessments intended to measure basic reading skills. According to Minnich,

early results of the pilot show smaller gaps in achievement between White students and students of color than on the state's current statewide assessments.

"So from my perspective," Minnich said, "the easy, low-hanging fruit is to do some of these things that are fairly straightforward. Find out what kids are being taught in the state and make sure those are the topics on the state assessments. Give kids choice about topics. Those are things that are fairly straightforward to do and they're things we largely don't do in assessments right now."

Randall, whose work is focused on making assessments anti-racist, stated that "It's not enough to keep developing an assessment that is not racist, and right now that has been our goal. We need to develop assessments that are explicitly anti-racist." Therefore, assessment vendors need to be constantly asking themselves "What groups are being privileged by this assessment? How and why? Who is being harmed? How do we disrupt that harm?"

"Each day in my work I ask myself what would it look like to develop an assessment system that centered students of color, not an assessment system that would give students of color access to whiteness?" Randall said. "That's how we address these issues related to bias."

"Each day in my work I ask myself what would it look like to develop an assessment system that centered students of color, not an assessment system that would give students of color access to whiteness."

*— Jennifer Randall
University of Mass. Amherst*



Confusion About the Role of Accountability

Lovell opined that much of the opposition to accountability is based on past rather than present federal policy, saying that when it comes to federal accountability, we “need to separate fact from myth. Federal policy around accountability completely changed in 2015. And I feel like there’s still a lot of PTSD around NCLB.”

Lovell noted that while under NCLB, schools faced specific and, according to some interpretations, harsh consequences for not making progress, “[under ESSA,] if a group of students consistently underperforms [schools] may be identified for support. ...I think if people had a better understanding that schools can actually get more money based on how their students are performing, that’s a logical system. But people think that if they do poorly on a test, they’re going to have money taken away. And that’s 100% not part of the federal policy.”

“People think that if they do poorly on a test, they’re going to have money taken away. And that’s 100% not part of the federal policy.”
—Phillip Lovell
All4Ed

Kubatzky agreed, saying, “It’s not to punish schools or point out where educators are doing poorly, other than get support to those schools and districts that need additional help.” While Garza noted that the idea that “teachers are going to get punished or fired they’re going to be evaluated based on the summative assessment—that’s not a part of ESSA. In fact ESSA prohibits the Department of Education from prescribing any part of a state’s teacher evaluation system.

“Another example,” Garza said, “is that tests are used to hold kids back. Again that is a state decision, and the Department of Education also has guidance that discourages states from using assessments in any way that punishes students.”

Broken Promises of Accountability Systems

Yet Randall made clear the consequences of assessments go beyond formal accountability systems, especially for communities of color. “In the communities where I work, accountability means you get these test scores back, they’re published in the Boston Globe, and everyone says ‘your kid goes to a crappy school and all the teachers there suck.’ ...So I do think that the way that communities of color have been dehumanized and brutalized by assessments, even since NCLB, is really quite real.”

Feliz struck a similar tone saying that we do need to acknowledge the history of racial discrimination in education, including assessment, such as when SAT and IQ tests were used as “evidence” of Black students’ intellectual inferiority. But, ultimately, she says statewide assessments are a critical tool in continuing to improve our schools:

“Standardized statewide tests, by shining a light on academic gaps, are intended to promote equity and hold a system accountable for educating all students...As a parent and a community member I am trusting my public school system to do right by my child, and if history is precedent then I have to do my part to hold my school system accountable. And there's still a little mistrust in our communities when it comes to not just testing, but the government's role in promoting discriminatory policies that disadvantage students of color. ...I think we've made a lot of progress and we just need to keep moving towards advancing equity in our education system.”

“Standardized statewide tests, by shining a light on academic gaps, are intended to promote equity and hold a system accountable for educating all students.”

*—Susie Feliz
National Urban League*

And Jimenez added that the prospect of additional federal resources is likely viewed as a broken promise by schools. “While schools are supposed to get additional money if they aren't meeting benchmarks, that money often doesn't materialize because there isn't enough to go around...While there's the intent in the law to provide a level playing field, the playing field is so uneven the federal resources can't get it all caught up.”

Misuses and Reactions to Accountability

A number of our panelists also suggested that how states and districts are implementing assessments and accountability—rather than the design of systems themselves—is causing problems and driving opposition to assessment.

For example, Lovell said that many states do not comply with federal requirements to report data and identify schools and, as such, are limiting the number and types of supports that could be available to schools, as outlined in a recent [Data Quality Campaign](#) report. “One of the bare minimum requirements of the law is that if students of color, low income kids, students with disabilities, if these groups of kids consistently underperform, the school is identified for support so that something can happen for those kids.

“And in one state they identified one school for having consistently underperforming Latino students, two schools for consistently underperforming African-American students, and this state has some of the largest achievement gaps in the country. That's just nonsensical. We need to implement the law with the eye of doing what's best for kids.”

Rodrigues said that schools are often creating an environment of anxiety around testing that is totally unnecessary. “I always hear that the test is so stressful. It stresses out our kids. It stresses out our teachers. It stresses out the school. But who's creating that stress? Who's creating the environment where the test is a life or death, anxiety ridden exercise?”

"If every time I take my kids to the pediatrician for a check up I get them all freaked out by saying 'oh my god they're going to take your temperature and if you have a fever that might mean you're sick,' then of course my kids are going to be filled with anxiety. So that tells me that the messaging that we as adults are giving is wrong."

Marshall also noted that ESSA only prescribes the end of the year assessments, not the other interim and diagnostic assessments administered by districts. "So this is a decision made at the district level about how much assessments are necessary. You definitely need some kind of a baseline but you really should not be doing this kind of wholesale every three weeks where you're breaking down instruction and not teaching." And it's this practice that likely leads to the perception overtesting—not the federally required assessments.

Finally, Rodrigues mentioned a perennial complaint about summative assessment data: that it takes far too long for districts and parents to get results. However, she said, the pandemic may have marked an end to this (NPU's [EPIC campaign](#) is focused on pushing for transformational change using data and federal pandemic relief). "We had springtime assessments and got that data to districts by July 1st here in Massachusetts. ...So they can't fool us anymore and say it's impossible to get this data or it can't be there fast enough. We can do this. We can figure this out. This is one of the silver linings from the pandemic that I hope actually stays with us."

"They can't fool us anymore and say it's impossible to get this data or it can't be there fast enough. We can do this. We can figure this out. This is one of the silver linings from the pandemic that I hope actually stays with us."

*—Keri Rodrigues
National Parents Union*

Watch recordings of each of the three sessions [here](#), [here](#), and [here](#).

