Modernizing Measurement:

A Policy Primer for the Next Generation of Assessment

Foreword by
Richard Culatta
Penny Schwinn

By Lauren Camera, former *U.S. News & World Report* Writer

Commissioned by

AMIRA LEARNING

Produced by

W/A

Foreword

The recent offer from the U.S. Department of Education for states to reimagine key tenets of the federal K-12 education law, the Every Student Succeeds Act, is a clear signal that it's time for states to tackle something that is long overdue: a genuine rethinking of assessment systems that have frustrated educators, failed students, and consumed billions of dollars without meaningfully improving learning. For those of us who have spent careers leading state and federal education systems, that signal feels like a turning point worth seizing.

We come to this issue from different vantage points—state system leadership and education innovation—but we share the same frustration. We have both watched well-intentioned reform efforts collide with the hard limits of federal mandates, immature technology, and a political environment that made meaningful innovation around assessment difficult to sustain. And we have both seen what happens when states respond to these constraints not by innovating, but instead allowing mediocre standardized tests and a dearth of formative assessments to persist while wishing for a different outcome.

Tennessee's testing history is a case study in the latter. Students and teachers spent years caught in a cycle of high-profile testing transitions—each one launched with genuine ambition, each one constrained by infrastructure that wasn't ready, procurement timelines that were too compressed, and platforms that buckled under pressure. It took years to process those lessons, rebuild trust with educators and parents and refocus the conversation on what information educators actually need to improve instruction—and how to get it to them in time to act on it.

At the federal level, multiple administrations have championed the shift from traditional, high-stakes testing to more innovative, technology-driven approaches and incentivized the development of next-generation assessments aligned to college- and career-ready standards. Leveraging technology to make assessments more responsive, accessible, and useful for both teachers and students did not falter due to lack of effort but lack of capacity to translate that innovation into widespread practice.

Both of our experiences point to the same conclusion: While the importance of assessment remains undeniable—ensuring all students, no matter their background, are making progress toward important academic benchmarks—the barriers to genuine assessment innovation have been real. Federal mandates constrained flexibility. Early online assessments were piloted before the technology was ready. Assessment tools couldn't reliably produce the kind of continuous, standards-aligned data that would make dynamic models viable at scale. And our implementation of them has been sloppy at best.

Summative assessment was never designed to drive instruction—which means the accountability calendar and the instructional calendar have always been in fundamental tension. It used to be that you had to optimize one or the other. Not anymore.

That's now changing. The device and connectivity gaps that once made online assessment unreliable are largely closed. Assessment technology has been transformed by AI and advances in psychometrics, enabling continuous measurement that adapts in real time, captures early literacy development even for the youngest learners, and aggregates into the kind of longitudinal data federal accountability requires—without the cost and disruption of a single high-stakes annual exam.

Perhaps most importantly, the policy environment has shifted. The Every Student Succeeds Act (ESSA) offers more flexibility than most states have already used–explicitly permitting, for example, multiple interim (formative) assessments to be used in place of a single annual test, provided they together produce valid, reliable, and transparent information about student learning. The Education Department's recent encouragement to state chiefs— backed by actual waiver approvals in Iowa and Lousiana—suggests the flexibility is now real, not theoretical.

The brief that follows makes the case for what is now within reach: a transition away from the end-of-year summative test—expensive, time-consuming, and instructionally inert—toward dynamic assessment models that generate continuous insight, inform teaching in time to make a difference, and ultimately accelerate learning for every student. The argument is not theoretical. States are already doing it. The evidence is mounting.

The door is open and the moment is now. We hope you'll walk through it with us.

Richard Culatta
CEO, ISTE+ASCD

Penny Schwinn
Former Tennessee Commissioner of Education

Table of Contents

Richard Culatta

Prior to joining ISTE+ASCD, Culatta served as the Chief Innovation Officer for the state of Rhode Island. In this role, he led an initiative to make Rhode Island the first state to offer computer science in every K-12 school and created a state vision for personalized learning. Culatta also served as the Executive Director of the Office of Educational Technology for the US Department of Education. In that capacity, he led efforts to expand connectivity to schools across the country, promote personalized learning and develop the National Education Technology Plan.

Penny Schwinn

Dr. Schwinn is a Senior Advisor at Whiteboard Advisors. She previously served as Tennessee education commissioner, chief deputy commissioner of the Texas Education Agency, and chief accountability and performance officer for the Delaware Department of Education. Following her time in Tennessee, Dr. Schwinn served in a senior capacity at the U.S. Department of Education, where she contributed to federal efforts to support states in accelerating student recovery and strengthening K-12 systems nationwide.

About the Author

Lauren Camera is a Senior Director at Whiteboard Advisors. She was a senior writer at U.S. News & World Report, best known for its comprehensive annual rankings of America's colleges and universities. For nearly two decades, she covered education policy and politics, detailing the education agendas of the last four administrations and providing expert analysis of the shifting landscape of child care, K-12 and higher education. Camera has also written for *Education Week*, *The Hechinger Report*, *Congressional Quarterly*, *Roll Call* and the *Chronicle of Higher Education*. She was a 2013 Spencer Education Fellow at Columbia University's School of Journalism and a 2020 finalist for the Society of Professional Journalists' award in the investigative category.

Acknowledgments

We are grateful to the experts who have contributed their time, energy and expertise to this policy paper. Their experiences attempting to develop assessments and craft policy that better supports teachers and improves student learning have charted a course that makes the promise of more dynamic assessment possible today.

Mark Angel
CEO and Co-founder, Amira Learning

David P. Cleary,
Principal, theGROUP; former Staff Director of the U.S. Senate Committee on Health, Education, Labor, and Pensions and Chief of Staff to U.S. Senator Lamar Alexander

Richard Culatta
CEO, ISTE+ASCD; former Chief Innovation Officer, Rhode Island; former Director, Office of Educational Technology, U.S. Department of Education

David DeSchryver
Senior Vice President and Co-Director of Research, Whiteboard Advisors

Chris Domaleski
Executive Director, Center for Assessment

Christina Grant
Executive Director, Center for Education Policy Research, Harvard University; former State Superintendent of Education, District of Columbia

Dianne Henderson
Vice President of Impact, Amira Learning

Catherine Hofmann
Senior Vice President for Government and Public Affairs, ACT

Tim Knowles
President, Carnegie Foundation for the Advancement of Teaching

Nat Malkus
Senior Fellow and the Deputy Director of Education Policy Studies, American Enterprise Institute

Ellen Dollarhide McCoy
Associate Director of the Center for Civics, Education, and Opportunity, Reagan Institute

Hillary Rinaldi
Vice President, Whiteboard Advisors

Roberto Rodriguez
Former Assistant U.S. Secretary of Education for Policy and Planning; former Deputy Assistant for Domestic Policy; former Chief Education Counsel to Sen. Edward M. Kennedy and the Senate Health, Education, Labor and Pensions Committee

Angela N. Romans
Executive Director, Innovation For Equity

Penny Schwinn
Former Commissioner of Education, Tennessee

David Steiner
Executive Director, Johns Hopkins Institute for Education Policy; former Commissioner of Education, New York State

Jane Swift
CEO, Education at Work; former Governor of Massachusetts

A Disclaimer

This primer doesn't provide an exhaustive history of assessment policy in the United States, nor was it intended to. It is also not without bias.

Amira Learning, a company engaged in the development of assessment tools, commissioned the report. At the same time, it was developed with a commitment to intellectual independence, rigor, and transparency in examining both the opportunities and the challenges of next-generation assessment.

To be sure, the rapid advancement of technology raises important and unresolved questions about equity, access, and the distribution of educational opportunity. While emerging tools hold promise for more responsive, personalized, and instructionally relevant assessment systems, there is always risk that new policies and approaches exacerbate existing disparities if systems—and practices are not thoughtfully designed and implemented. There are growing concerns about screen time and the use of technology in schools altogether. These tensions are real and consequential, and they warrant careful consideration from policymakers, educators, and stakeholders alike.

At Whiteboard Advisors, we approach this work with both optimism and caution. We believe in the transformative potential of technology to improve teaching and learning. At the same time, we recognize the responsibility to ensure that innovation leads to tangible improvements for all students—not just some—and that it supports, rather than supplants, the professional judgment of educators.

This report does not advocate for a single solution or prescribe a definitive path forward. Instead, its purpose is to provide a clear and grounded understanding of the legal landscape and recent policy history shaping K–12 assessment. By establishing a shared foundation of knowledge, we aim to support more informed, constructive dialogue.

Ultimately, meaningful progress will depend on the ability of diverse stakeholders to grapple with complexity, confront tradeoffs, and work collaboratively toward assessment systems that both advance academic outcomes and strengthen instructional practice.

Introduction

Hurricane Katrina barreled into Louisiana in 2005, flooding New Orleans, destroying homes, schools, and infrastructure, and displacing hundreds of thousands of residents. The storm caused widespread loss of life, economic ruin, and a collapse of public services—including a public education system that was already reeling from underinvestment and poor performance decades before the natural disaster.

As students and families fled the district—and state—in search of stability while first responders stitched Louisiana back together, school leaders grappled with how they could possibly rebuild a school system that met the requirements of the recently passed No Child Left Behind Act. That was until state and federal policymakers identified a little-known provision of the law that allowed the Secretary of Education to waive certain requirements in extraordinary circumstances.

In the end, the federal waivers that the Education Department granted to the state opened the door to the radical redesign of an entire school system—one that altered Louisiana's accountability system, allowed for alternate assessments, and fueled the reinvention of a district. Over the next five years, student scores grew between 11 and 16 percentiles compared with similar students and districts, with high school graduation, college entry, and college graduation rates improving as well.

Today, we're in the midst of a different sort of crisis that is revealing the limitations of federal mandates that undermine state and local ingenuity. National data shows math and reading scores on a downward spiral and achievement gaps widening —a post-pandemic quagmire that billions of dollars in long-spent federal recovery funds barely budged. The most recent results from the National Assessment of Educational Progress show high schoolers performing at lower levels than they did two decades ago in reading and math.

Meanwhile, rapidly shifting workforce demands, driven by the proliferation of artificial intelligence and other emerging technologies, are challenging our understanding of what our students will need to know to be successful in life. The cumulative impact of sinking test scores and the transformation of our economy has state education leaders scrambling to rethink accountability and assessment in a way that delivers teachers the types of insights needed to drive academic improvements, identify and interrupt gaps in learning and equip students with a whole new set of skills.

At a moment when the nation faces not just one, but multiple education crises there is an opportunity for state leaders to rethink accountability, redesign assessments, and pursue innovations their students urgently need.

Many states are beginning to move in this direction, and the promising innovation underway can provide for more dynamic assessments that measure not just student achievement, growth, and mastery of content at grade level, but provide the sort of insight for educators (and policymakers) that can be used to improve outcomes.

Too Many, Too Little, Too Late

It's not a novel recognition that rear-facing summative tests are, alone, ill-equipped to drive the type of urgent academic improvements we need: A static achievement metric tells us little about student growth and isn't helpful in informing educator instruction. The tests themselves are massive line-items on state budgets and eat into valuable classroom learning time.

Each year, dozens of bills are introduced in blue and red states alike with the goal of reducing the amount of time that children spend testing and that teachers spend preparing them for those tests. And yet for more than two decades, we've witnessed a sort of assessment "arms race."

The sort of one-time, high-stakes summative measures, often used by states, provide information too late to be instructionally useful and have fueled demand for formative measures that provide information that can inform district and classroom decisionmaking. Interventions, which yield increasingly granular and accurate data, and benchmark assessments provide yet another layer of decision support in an era where the practice and process of education are increasingly digitized.

Historically, there have been compelling legal and instructional reasons why both summative and formative assessments were necessary. But a new generation of assessments that fulfill both purposes are beginning to upend conventional thinking.

State and federal policymakers have made (at times, bold) attempts to make the move toward more dynamic assessments in the past—ones that continually evaluate students without interrupting daily instruction, equip educators with real-time feedback to make lessons more effective, and map directly onto state standards to ensure accountability and equity requirements. But technological capabilities and public policy have never been aligned like they are today.

While our ambitions haven't changed, the environment we're operating in has. We are now in an era defined by the increasingly widespread availability of broadband and mobile computing, along with significant psychometric and technological advances that simply did not exist as recently as 2008, when the federal Race to the Top competition marked the last major federal push for assessment reform.

> **Technological infrastructure is more robust:** There is ubiquitous use of laptops and devices by students. Schools are more connected to broadband than ever before, and those connections are stronger than they've ever been.
>
> **The policy environment is more flexible:** The restrictive NCLB years gave way to the Obama-era of competitive grants and waivers, which in turn gave way to the flexibility built into ESSA. Today, the current administration is offering even more latitude to allow for state and local innovation.
>
> **Assessments technology is more advanced:** From advanced psychometrics to speech recognition technology and AI integration, assessments are more powerful and capable of providing reliable, precise

insights into student learning and growth while preserving the sort of integrity necessary for longitudinal measurement and rigorous accountability.

Research is more precise: Breakthroughs in assessment research and in the science of learning suggest better ways to collect and make use of real-time data on student learning and achievement in order to tailor and personalize instruction to accelerate progress without legacy concerns of "gaming," which some critics argued would render everything high stakes.

Taken together, these advances open the door for states and districts to adopt assessments that constantly gauge student growth and achievement, amassing enough insight to satisfy federal requirements over time without the cost and time necessary for annual, summative testing.

A handful of states are already experimenting with the idea: Maine administers assessments in the fall and spring (with an optional winter administration), generating math and reading growth scores at each interval to give teachers timely, actionable data on student learning. At year's end, a summative score is derived from the spring assessment, while scores from earlier administrations allow educators to measure growth both within and across school years. Delaware is testing out something similar, with students taking three formative assessments in social studies, which are then aggregated to produce a summative score. Both examples eliminate the need to administer end-of-year exams. Similar efforts are underway in Louisiana, Montana, North Carolina and other states.

In September 2025, in a significant move toward dynamic testing, Texas Gov. Greg Abbott signed a bill replacing the state's single end-of-year summative test with a model that assesses students at the beginning, middle and end of the year, providing timely feedback to inform instruction and offering students multiple opportunities to show what they've learned.

Why aren't more states moving in this direction?

How Did We Get Here?

The No Child Left Behind Act of 2001 required states, for the first time, to administer annual assessments in reading and math to all students in grades three through eight and once in high school, in grades 10 to 12. Schools were held accountable for making adequate yearly progress (AYP), by which a certain percentage of students had to be proficient in reading and math.

The percentage of students required to meet AYP increased each year until the 2013-14 school year, when, aspirationally, all students, including those with learning disabilities and English Language learners would be on grade level. (Of course this would never happen.) Schools that did not meet AYP faced a set of increasingly punitive consequences for each year they missed the mark.

Though brokered in a bipartisan spirit of wanting to increase student achievement, the law quickly devolved as states "gamed cut scores"—setting proficiency bars artificially low to show progress—and testing became a central tool for federal oversight, teacher evaluation, and school ratings. Albeit well intentioned, the years that followed (notably, absent reauthorization, which could have afforded the opportunity to course correct and tweak the law) were often characterized as a "race to the bottom."

To be sure, there were notable and, in hindsight, somewhat durable areas where educational outcomes improved. But state political leaders gamed the system in certain cases and gaps widened between state scores and NAEP, with little or no way to compare performance across states without secondary analysis. The resulting pushback on accountability, no doubt, set progress back.

Utah was one of the most vocal critics of the dynamic, petitioning the Education Department in 2007 for a waiver to use its own state-developed testing system—a formative model that measured student growth instead of a static proficiency score. The department denied the waiver, further fueling resentment over federal control among states eager to adopt new assessment and accountability models that sought to measure student progress with growth models.

By 2011, the Education Department began formalizing a policy of offering waivers from NCLB's AYP provision in exchange for implementing new assessments aligned to new standards—at least for those who hadn't already done so through Race to the Top.

In a watershed moment for state flexibility from NCLB and a sign of what was to come for testing in the U.S., the Education Department began granting those waivers in 2012—nearly 40 in all. While few states proposed sweeping changes to assessment, a handful, like Idaho, proposed replacing AYP with a five-star rating system that graded schools based on student growth and proficiency on state tests, graduation rates, and career- and postsecondary-readiness indicators.

In 2015, Congress and the White House worked in a bipartisan fashion to finally replace NCLB

2008 | A Bold Federal Experiment

In 2008, the Obama administration launched a $4.35 billion federal grant competition called Race to the Top, which directed funding to states that promised to overhaul their education systems in very specific ways—including by developing common assessments.

In doing so, more than 40 states and the District of Columbia joined one of two federally-funded testing consortia—PARCC (Partnership for Assessment of Readiness for College and Careers) and Smarter Balanced Assessment Consortium—creating new standardized tests.

The goal was to raise the quality, consistency, and usefulness of state testing systems. Both consortia were designed as online, adaptive assessments—with the goal of measuring student ability more precisely and returning faster results that could deliver actionable insights, not just end-of-year scores.

But these grand testing experiments were largely viewed as failures, with national headlines bemoaning sinking test scores and massive implementation failures.

Some states administered two sets of tests in one year—their old, individual state tests and whichever consortia test the state had adopted—prompting concerns from teachers that test prep dominated classrooms and fueling an "opt-out" movement among parents who refused to let their kids take new exams.

There were also implementation and scoring challenges: The tests were computer-based, but many schools, especially in rural areas or low-income neighborhoods, lacked the technological infrastructure to administer them effectively. Early roll-outs experienced system-wide crashes, long testing times and general confusion. Tests were harder and so scores were lower—further alarming parents and vexing policymakers trying to input them into new teacher evaluation systems that dictated things like salary.

with the Every Student Succeeds Act. In doing so, it enshrined the testing requirements in grades three to eight and once in high school, but removed the AYP requirements. It also gave states more flexibility over things like how test scores are used in accountability systems and how much weight they carry in school ratings. The law also allows states to replace their state high school exam with nationally recognized tests like the SAT and ACT, and allows up to seven states to experiment with alternative innovative assessments systems.

What Federal Law Actually Does (and Doesn't) Require

The Every Student Succeeds Act still requires annual statewide testing in reading/language arts and mathematics for grades three through eight and at least once in high school. Science must be tested at least once in elementary, middle and high school.

Legislative Text on Assessment

(I) in the case of mathematics and reading or language arts, be administered—

(aa) in each of grades 3 through 8; and

(bb) at least once in grades 9 through 12;

(II) in case of science, be administered not less than one time during—

(aa) grades 3 through 5;

(bb) grades 6 through 9; and

(cc) grades 10 through 12; and

Tests must be aligned to state content and achievement standards and results must be disaggregated by student subgroups. States also must design systems to identify and support underperforming schools using multiple measures, of which test scores are one component.

Notably, the law dictates that at the state's discretion, the required test can be administered through a single summative assessment or through multiple statewide interim assessments during the course of the academic year that results in a single summative score that provides "valid, reliable and transparent information about student achievement or growth."

While about a dozen states are experimenting with administering multiple interim assessments as opposed to a single summative one, no states currently take advantage of that option at scale.

Legislative Text on Flexibility

(viii) at the States's discretion

(I) be administered through a single summative assessment; or

(II) be administered through multiple statewide interim assessments during the course of the academic year that result in a single summative score that provides valid, reliable, and transparent information on student achievement or growth;

Notably, federal law does not:

- Dictate whether states focus on student achievement or growth[1].
- Define what type of assessments states choose, be it formative or summative.
- Define the type of formative assessments states choose, if they choose that model.
- Define how states collect data from formative assessments.
- Define how states weigh that data, so long as it provides "valid, reliable and transparent" information.

[1] *ESSA requires states to use academic achievement measures (proficiency rates) in their accountability systems. States cannot exclusively use growth; proficiency on state assessments.*

The Future of Assessment is Now

In the 25 years since NCLB, the fields of edtech and psychometrics have undergone a fundamental transformation. Assessments that once required lengthy, disruptive testing windows can now be embedded into daily instruction. Interventions now yield more precise and timely insights into student learning, and digital tools and learning platforms have made frequent, low-stakes assessments easier to administer and analyze. These allow for continuous, precise insights into student learning without pulling teachers out of teaching or students out of learning. The question is no longer whether better tools exist—it is whether states will use them.

The pandemic's forced investments in technology also fueled a step change in the technological capabilities of our schools. Today, learning platforms, apps and educational software are more reliable, intuitive, and capable of capturing detailed data about student learning than ever before—and they're everywhere. They blend diagnostic functions, with decision support for educators and, in many cases, prescribe recommended pathways and interventions to improve student outcomes—sometimes, in close to real time.

Digital formative measures that check in on what students know at the beginning, middle and end of the year—as well as monthly, weekly, sometimes even daily—already exist.

Data generated through everyday instructional activity—reading logs, problem sets, writing assignments, digital learning platforms—can now be aggregated, synthesized, and reported in ways that satisfy federal accountability requirements without a separate, standalone test. The information teachers are already collecting about student progress and skill gaps can be structured to serve district, state, and federal decisionmakers simultaneously. Assessment and instruction no longer have to be separate events.

All of this progress and yet the vast majority of systems have layered new assessments on top of old ones rather than replacing them. Students now move through a calendar crowded with diagnostic, benchmark, interim, and summative tests—each designed for a different purpose, none designed to coexist. The result is more time spent on assessment than at any point in the last two decades, with no corresponding improvement in the quality or timeliness of the information produced.

The good news is that while federal law requires specific testing in specific grades, it also allows for significant flexibility. Like never before, ESSA opened the door to a new era of dynamic testing that honors critical accountability priorities, with instructional relevance and—above all—a dramatic reduction in the time and money spent on tests that provide precious little information, far too late to do anything about the problems they identify.

Some states are already experimenting with new forms of assessment. But far too few have taken advantage of the opportunity.

Looking ahead, these dynamic models will increasingly become computer-adaptive, adjusting

question difficulty based on student responses. They'll measure both content mastery and growth, giving more precise data about what students actually know.

Artificial intelligence has extended those capabilities further and faster than most policymakers have registered. AI-assisted scoring can now evaluate oral reading fluency, open-ended written response, and complex problem-solving tasks at scale—delivering results in hours rather than months, at a fraction of the cost of human scoring. Gone are the days when human scoring was considered the gold-standard in standardized testing. Advances in technology are reducing human errors and boosting reliability. Most importantly, this moves resources into classrooms.

AI can also convert that performance data into individualized, actionable feedback for teachers and students in time to change what happens next in the classroom. That feedback loop—precise, timely, and instructionally relevant—is what the current model has never been able to provide. It is now within reach for every state ready to ask for it.

None of this is to say that a summative, nationally-benchmarked test like NAEP isn't necessary. It holds tremendous value because it serves as the only consistent, objective measure of student achievement across the United States. It allows for vital comparisons across states and districts and over time, and it provides a high-standard benchmark for proficiency. That standardized benchmark helps policymakers and educators identify learning gaps and evaluate educational progress in grades 4, 8, and 12.

Moreover, the role of NAEP as the "Nation's Report Card" is critical to serve as an independent audit of sorts to the quality and integrity of state assessment systems. Just as a doctor can measure a patient's cardiovascular fitness whether the patient climbs ten flights of stairs, jogs three miles or runs a set of wind sprints, so too do education researchers have the psychometric capabilities today to crosswalk unique state assessments toward a well-designed, high-quality independent assessment.

Assessment Types in Practice

Diagnostic Assessment	Formative Assessment	Interim / Benchmark Assessment	Summative Assessment
PURPOSE			
Identify which specific skills a student has mastered, which they are ready to learn next, and which foundational predecessor skills still need to be addressed—enabling targeted instructional decisions	Generate immediate evidence of student learning within and across lessons to inform instruction in real time	Provide medium-term evidence of student progress toward grade-level learning goals across a unit or marking period	Verify whether students have met measured grade-level standards at the end of a defined instructional period
UTILITY			
High utility to students, teachers, parents, and administrators—provides actionable roadmap for each learner	High utility to students and teachers; supports immediate instructional adjustments	High utility to students, teachers, parents, and administrators	Low utility to students, teachers, parents, and administrators—but plays an outsized role due to state and federal accountability requirement
PRIMARY USE			
Instructional—establishes where each student is relative to grade-level and/or intervention	Instructional—informs what happens next in the lesson, unit, or instructional sequence	Instructional, evaluative, and/or predictive—measures whether students are on track toward proficiency	Evaluative—determines whether students have met standards; feeds state and federal accountability systems
TIMING			
Initiated at or before the start of instruction to establish a baseline; Administered at select intervals during the school year	Administered within or between class periods; ongoing throughout the instructional cycle	Administered at select intervals during the school year—typically 2–4 times	Administered at the end of a unit, course, or school year
RESULTS AVAILABLE			
Immediately—within minutes or hours of administration	Immediately—within minutes or hours of administration	In enough time to affect instruction within the course or marking period	As required by school, district, or state policy—typically weeks or months after administration
EXAMPLES			
Adaptive (computer or multi-stage) assessments, reading survey, math inventory tests, running records	Exit tickets, quizzes, observation, classroom discussion, writing assignments	Fixed periodic assessments; evaluating students against grade-level standards at defined points in the year	End-of-course exams, state summative assessments, college entrance exams

To be sure, formative and summative assessments both can be designed as standardized or non-standardized tools depending on their purpose. Well-designed standardized formative assessments, in particular, can provide valuable, comparable data across classrooms and states while remaining low-stakes—informing instruction and progress without being used for consequential decisions like teacher pay or college admissions.

Conclusion

Current summative models are expensive, lengthy and antiquated. Artificial multiple choice questions push kids to guess while prep time and the test itself cost educators valuable hours of instruction. Replacing this structure with multi-layered measurements—culled through in-class exercises and natural conversations with teachers and peers—provides a more accurate, equitable and instructional useful evaluation of student achievement.

It's imperative to understand how advances in technology have unleashed the power of dynamic assessments to have better observational abilities, deep-learning statistical models that go beyond simple aggregation and capabilities to adapt in real-time to student behaviors. A single test at a single moment is too subject to error, too static of a data point and simply too primitive.

Today's technology allows us to think outside the box to change the structure of assessment, which has been left virtually untouched for three decades. And while assessing students in K-3 is much different than assessing those in middle and high school—the latter requiring more standardization—dynamic assessments will unlock our ability to create longitudinal tracking systems for pre-K-12 with ease.

We're moving from a moment when assessments provided educators little to no sense of student progress, to one in which the data flows continuously through tasks that are mapped to standards.

Importantly, as states increasingly move toward dynamic assessments, successful implementation will depend in large part on whether they engage parents thoughtfully and help them understand what their child's results mean. This has historically not been the case, as evidenced by a decades-long "perception gap"—the difference between parents' perception of their child's achievement and the reality. The challenge has been ensuring parents receive clear, timely information—and that results are presented in an accessible way that includes concrete next steps.

This outreach and partnership building will be particularly important during a moment of increasing skepticism among parents to technology in schools. With some research showing that testing mode matters in relation to how students learn, it should go without saying that efforts to rid schools of technology will undercut efforts to modernize assessment.

Curriculum-aligned assessments will play a critical role, too. Tying what students are assessed on to the materials they have studied and administering that assessment across multiple windows during the year ensures we safeguard against measuring knowledge gained through out-of-school experience and thus socioeconomic status. After all, when the

test bears no relationship to what students have been taught, it makes measuring learning less clear and concrete, and fails to reinforce the importance of studying the curriculum itself.

This is an important moment for assessments—marking a well-mapped shift from rear-view mirror testing that provides little more than an autopsy of student achievement to a system that actively monitors and provides real-time feedback. Once defined by students filling in bubbles on a paper test or computer screen, this powerful and accurate new model is defined by continuous observation.

The current approach to K–12 assessment is falling short for every stakeholder. While legacy systems may be psychometrically sound for high-level oversight, they lack the instructional sensitivity required to provide the timely, actionable insights needed to improve student learning. Latency undermines equity aspirations that often undergird accountability law and policy. And because these measures are designed for broad accountability rather than daily classroom utility, they often leave teachers discouraged, heighten anxiety for students, and offer parents an incomplete or confusing picture of their child's progress—all while giving policymakers information that is difficult to use well. Continuing to rely on this outdated model limits our ability to support students effectively.

Can dynamic assessment offer a path forward?

This paper was designed to provide an understanding of the history, policy, impediments, and attempts to improve educational assessment at scale. Want to learn more about what dynamic assessment could look like in practice? Stay-tuned, for Part II, which takes a deeper look at what this would look like in practice and a working definition for an emergent category of "dynamic assessment."