



# Now is the Time to Reimagine Assessments

Arthur VanderVeen, Ph.D.  
President and Chief Executive Officer



The pandemic has upended business as usual and forced all of us to rethink fundamental assumptions about how we educate our students. This is a good time to remember that with every challenge comes opportunity. With federal funding flowing in abundance and states and districts ready to make major changes, we have the chance to boldly re-imagine our educational models—including assessment—to create systems that increase equity, promote deeper learning, and embrace a more holistic view of education.

## Seizing Opportunity

As this pandemic stretches on, we are seeing data that shows the impacts on student learning are unprecedented. For many students, the impacts can now realistically be called “lost learning,” as their schools simply are not equipped to address so much developmental delay. Second graders who cannot name their letters—something expected of kindergartners—will face enormous academic challenges for years to come. And we know these impacts are not distributed evenly; the best data we have—data from states that administered their spring 2021 statewide assessments—shows that historically underserved students (economically disadvantaged students, students of color, English language learners, and students with special needs) experienced greater learning loss than white students. We know that because the state assessments provide the most valid and reliable longitudinal measure of student learning that can be compared year over year, across schools and student groups. The data are unequivocal: our best efforts to respond to the pandemic have not met the challenge. The inequities of our educational system have been laid bare by this calamity; now is the time to address them.

This includes addressing the limitations of our assessment systems. No state administered its summative assessment in 2020, and many states did not administer in 2021. In an unexpected way, these cancellations have created an opportunity to rethink our assessment and accountability systems.

The two-year lapse in normal testing caused by the pandemic will make it impossible for states to calculate the growth component of their ESSA plan. States will not be able to calculate growth spanning from the spring 2019 test administration to the next administration in spring 2022 in ways that reliably support their efforts to evaluate schools and use data to direct resources and support. The spring 2022 administration will establish a new baseline from which future growth trend lines will be calculated.

This break in trend represents an opportunity to rethink our assessment and accountability models and to rebuild them in ways that correct for their limitations. We have an opportunity to reexamine the overall paradigm that has led to very little innovation in summative test designs, an opportunity to rethink assessment as a means, not an end: A means for expanding equity and closing opportunity gaps; a means for inspiring student engagement, not stifling it; a means for deepening practice in the classroom; and a means for promoting a more holistic view of student learning.

**We have an opportunity to rethink assessment as a means, not an end.**

## Re-imagining Assessment

High-quality statewide assessments provide key stakeholders in the educational system with comparable student data, enabling insights critical to best understand how students are performing over time across states and school districts. This data enables policymakers, state and district leaders, and educators to identify inequities and allocate funding, resources, and support to historically underserved students and those more significantly impacted by the pandemic. More broadly, an assessment system that supports the public reporting of disaggregated data by student group, and the ability to compare districts and schools allows the public to understand how the system is serving all students.

That said, the push to reimagine next-generation assessments derives in great part, I believe, from our having narrowed our evaluation of student learning to a single summative score. While a single summative score effectively serves accountability purposes, it should be viewed in the broader context of other evidence that converges toward a richer interpretation of student development or readiness. We need to conceive new ways to broaden our measures of student learning while also maintaining the ability to hold schools accountable to serving all students equitably. That's not an easy task and it will require thought, time, and investment.

## Five Design Principles

A process for re-visioning assessments that meets these objectives should start with these five design principles:



**Design for coherence**



**Promote equity**



**Consider new holistic measures of student learning**



**Encourage student agency**



**Design scoring and reporting for coherence and usability**

### 1 Design for coherence across state, district, and classroom assessment models.

Our current systems of assessment are out of balance, incoherent, and harming kids. We have layered on a combination of mis-aligned statewide accountability assessments, district growth assessments, and local instructional assessments that collectively provide incoherent feedback on student learning and have overburdened educators and students with too much testing time. The pandemic-driven disruptions to in-person schooling have made instructional time more precious than ever; we can no longer afford as a nation to impose this muddle of mis-aligned assessments that make it difficult for schools to respond effectively to unprecedented learning loss.

Rather, we need balanced, coherent, and streamlined systems of assessment that complement one another, minimize overall testing time, and optimize the value of information used to support student learning and development. We need assessments that align to and work with local curriculum and instruction to accelerate student learning.

Researchers in the learning and measurement sciences have been promoting a vision for coherent systems of assessments for some time. A seminal National Research Council committee report, [Knowing What Students Know](#)<sup>1</sup> provided a roadmap for how states could develop coordinated systems of assessments that work together at the state, district, and classroom level to reinforce a shared set of learning goals. According to Pellegrino and his colleagues, “vertical coherence” across levels of the system derives from an explicit, common model of how students learn—how they build and represent knowledge and increasingly apply their knowledge in critical thinking and problem solving. A clear model of how students build and apply content knowledge lays the foundations for coherence across levels of the system.

<sup>1</sup> Pellegrino, Chudowsky, & Glaser, 2001

How is it that we have allowed our systems of assessment to remain so imbalanced for these 20 years? ESSA now allows and even encourages states to design balanced assessment systems that could include competency-based assessments, curriculum-embedded assessments, interim assessments, or performance-based assessments that foster deeper learning. But with minimal policy flexibility, no technical assistance, and limited funding, states have struggled to design innovative, comprehensive systems of assessment.

States have focused exclusively on the state accountability assessment, which is (appropriately) designed to measure end-of-year standards mastery. States rarely provide learning development frameworks that guide districts in the design or selection of local curricula and assessments that align to a common vision for how students progress towards end-of-year standards mastery. And, with limited resources, districts purchase a patchwork of commercial, off-the-shelf, interim assessments that have not been designed to align to an explicit model for how students learn through the course of the year.

**To achieve vertical coherence across state, district, and classroom assessments—and horizontal coherence with selected curricula—states and districts should partner on a process of developing learning development frameworks for each grade or grade span that articulates how students develop through the course of the year toward end-of-year standards mastery.**

To achieve vertical coherence across state, district, and classroom assessments—and horizontal coherence with selected curricula—states and districts should partner on a process of developing learning development frameworks for each grade or grade span that articulates how students develop through the course of the year toward end-of-year standards mastery. With such a framework, districts can develop or select interim/formative assessments, curricula, and instructional materials that align to and reinforce their model of how students learn. Reporting assessment scores in context of such a learning development framework also enables teachers, students, and families to meaningfully interpret test scores in ways that can be directly tied to instructional decisions. This all supports better learning opportunities for students.

New Meridian is investing in R&D to build a new system of through-course testlets aligned to research-based learning progressions that articulate how students develop from initial understanding to end-of-year standards mastery. A set of short testlets for each grade level report where students are on the learning progressions so that teachers and students can direct next steps. The testlets can be flexibly aligned to local curricula to measure students' developing knowledge and skills close to the moment of instruction and in context of what is taught in the classroom. Finally, we plan to conduct research to validate that testlet scores will aggregate over the course of the year to a summative measure of standards mastery.





New Meridian's system of testlets thus accelerates states' progress toward more balanced systems of assessment:

- Supports vertical and horizontal coherence by aligning classroom, interim, and state summative assessments to a common model of how students learn throughout the year.
- Provides timely and meaningful reporting of student progress against explicit learning progressions, improving instructional decisions and interventions.
- Directly measures the diverse concepts and skills outlined in the learning progressions, supporting a broad and rich content-based curriculum.
- Reduces overall testing time by aggregating through-course assessments to a summative score.
- Increases equity by assessing student learning in context of what is taught in the classroom, leveling the playing field among students with diverse and unequal social and cultural life experiences.

## 2 Maintain an overarching commitment to equity through comparability.

NCLB's greatest contribution in moving toward a more equitable public education system was its requirement for transparency in how well schools are serving all students. State summative assessments made that possible by providing a reliable and comparable measure of how well students learned the state's grade-level learning standards, a measure that could be disaggregated across districts, schools, and student groups. This allowed stakeholders—including families—to compare how well their local school was serving their child, irrespective of where they lived, their race or ethnicity, or whether their child had special needs or was an English learner. Parents finally had an objective, comparable and fair measure of their children's learning progress that wasn't subject to the implicit biases of either low expectations or inflated grades, both of which can harm a student's future potential.

**In our enthusiasm to reimagine next-generation assessments, we cannot weaken our commitment to reliability, comparability, and fairness. Next-generation assessment systems cannot fail to provide a reliable, comparable measure to support our need for transparency into how schools are serving all students.**

In our enthusiasm to reimagine next-generation assessments, we cannot weaken our commitment to reliability, comparability, and fairness. Next-generation assessment systems cannot fail to provide a reliable, comparable measure to support our need for transparency into how schools are serving all students.

The simplest and most direct way to do this is to maintain some form of summative assessment that efficiently meets requirements for reliability, comparability, and fairness and provides data that can be disaggregated to support transparency and accountability for school performance. If we assume other measures will broaden how we assess the richer dimensionality of student learning we value, we can shorten the summative test by focusing on priority standards. This, however, would require greater flexibility from the U.S. Education Department.

A number of states have expressed interest in eventually doing away entirely with an end-of-year summative test and aggregating several through-course tests to provide a summative score. Through-course assessments increase the instructional value of test data that is available in time for educators to adjust curriculum and instruction. There is great promise in this approach. Just as this is a more holistic representation of how students think and learn, we must now consider how assessments can more holistically measure student knowledge and skills in ways that are valid and coherent.

### 3 Promote and research more holistic views of student learning.

Once we accept the idea that multiple, complementary and aligned assessments are needed to meet the variety of purposes for which we test, and that common assessment frameworks based on a shared model of student learning can provide vertical and horizontal coherence across the system, we have the foundations for researching how to integrate a variety of measures into our system.



**Classroom formative and district assessments.** The state's assessment frameworks should provide guidance for how local formative and interim assessments measure the more fine-grained knowledge, skills, and discipline-specific practices to inform instruction. These assessments should also be aligned with the state assessment framework so that finer-grained measures of student learning can be interpreted in the context of larger summative claims.



**Classroom assignment grades.** Academic success derives from much more than just content knowledge and skills. It includes critical habits of mind, study skills, motivation, and persistence—all of which are more accurately reflected in classroom grades than on a single, end-of-year test. Assignments can be tagged to the state assessment framework so that grades contribute to a cumulative estimate of how students are learning and developing mastery in each subject domain. Statistical analyses must be conducted, however, to identify where implicit bias, systemic differences in expectations for student potential, or other factors cause local grading to deviate from other external measures (like the summative test). These analyses can disclose where systemic bias may need to be addressed.



**Graduate profiles, content modules, and portfolio defenses.** Graduate profiles and portfolio defenses typically articulate a more holistic set of 21st-century skills that students should develop, including critical thinking, creativity, collaboration, and communication.

**Performance tasks** ask students to frame a problem, conduct research, analyze information, work across disciplines, and develop solutions. They also ask students to plan, collaborate, and communicate. Incorporating these performance-based measures will enable districts and states to broaden their claims about student learning and anchor their educational systems in a more holistic vision of student development.



**Culturally affirming assessments.** Little has been published on culturally affirming assessments, but the principles of culturally relevant education suggest that they provide opportunities for students to demonstrate mastery of content knowledge and discipline-specific practices in ways that are relevant to their culture and identity, thus **increasing engagement, motivation, and learning outcomes**. As a practice of empowerment, developing culturally affirming assessments would include involving local communities in test design and development, engaging diverse stakeholders to describe the learning progressions, achievement-level descriptors, and evidence statements in the assessment frameworks in culturally relevant ways.





#### 4 Allow for greater student agency.

Systems of assessment should provide positive opportunities for students to demonstrate their learning by creating space for more student choice. [Research](#) on student agency in assessment and learning suggests that empowering students to set goals, monitor progress, and adjust their learning strategies increases motivation and feelings of self-efficacy, which in turn leads to students investing more effort in their learning. There are several ways to incorporate these ideas into assessment systems:

- **Summative assessments could be designed to allow students to choose from a set of texts or tasks that schools have taught, so they can demonstrate their ability to read and analyze passages they are familiar with.** The simple act of providing students some choice from a set of assigned topics or texts helps compensate for differences in background knowledge, better aligns tests to what students have had an opportunity to learn, and would likely reduce testing anxiety because students would feel better prepared.
- **Provide on-demand, competency-based assessments** that enable students to set goals, choose from among a set of tasks to demonstrate mastery of learning objectives, and schedule testing based on readiness.

## 5 Develop score interpretation models and data systems that reinforce coherence and usability.

A system of assessments that incorporates multiple measures to develop a more holistic view of student learning needs a scoring model and reporting system that reinforces the above design principles and ensures the information is useful for intended purposes:



**Scoring and reporting to support coherence and usability.** Scoring and reporting should be differentiated to the information needs at each level of the system while reinforcing vertical coherence. Local assessment data (e.g., grades, through-course assessments, portfolios) should be mapped to the larger assessment framework to contextualize formative and through-course data within summative claims. This will help build the evidence base and thus the validity of claims about student mastery at all levels of the system. Stakeholders will gain confidence in the entire educational system when they see the relevance of assignments and assessments within a meaningful set of learning goals.



**Scoring and reporting to support comparability.** Combining multidimensional measures must be managed in order to be sufficiently reliable to support comparability. Calibrating performance-based assessment scoring requires clear descriptions of the underlying constructs being measured, clear rubrics and scored exemplars, training and calibration of scoring and grading, performance standards validation, and comparisons to the summative test. Research and evaluations of the ambitious New Hampshire PACE pilot suggest that developing a statewide system of local assessments that maintains comparability takes time and effort and poses logistical and technical challenges. To accommodate these challenges, educator training and involvement will be critical as we move to a more multi-dimensional approach. This training will increase teacher skill sets and elevate classroom practice as it helps create a new and more useful assessment system.

**Educator training and involvement will be critical as we move to a more multi-dimensional approach. This training will increase teacher skill sets and elevate classroom practice as it helps create a new and more useful assessment system.**

## Now is the Time

These five principles provide a framework for envisioning coherent next-generation assessment systems. They attempt to balance a desire to diversify our measures of student learning in ways that reflect the multidimensional, life-ready students we hope to develop, while also maintaining the comparability that is required to make sure our schools are supporting all students equitably. Those two goals—even if competing and technically at odds with one another—should drive our investments over the next few years in assessment and statistical research and design.

Meeting that challenge will require coordinated commitments across multiple stakeholders: the U.S. Education Department must consider the policy environment that would support this R&D (e.g., expanding and funding IADA pilots); philanthropy should fund research and technical assistance to support state and district staff (who are certainly already over extended) in this deep work; and states and districts planning their investments in recovery should collaborate to pilot and evaluate components of next-generation assessment systems in a process of continual improvement.

Crossing the chasm from narrowly focused single assessments to next-generation assessment systems will be challenging. But the pandemic has given us the opportunity to reset the rules, and America's commitment to recovery has provided the resources to make it possible.

Now is the time to try.

## ABOUT THE AUTHOR

**Arthur VanderVeen, Ph.D.**, is President and Chief Executive Officer at New Meridian, responsible for innovation, strategy, and customer engagement at the company he founded in 2016. Dr. VanderVeen's 20+ years of in-depth experience in education assessment and technology ranges from Compass Learning and the New York City Department of Education to the College Board.