



Design principles for new systems of assessment



R&D

Rather than being led by national testing mandates, state and local leaders should design balanced assessment systems guided by coherence, research on learning, and attention to equity.

R&D appears in each issue of *Kappan* with the assistance of the **Deans Alliance**, which is composed of the deans of the education schools/colleges at the following universities: George Washington University, Harvard University, Michigan State University, Northwestern University, Stanford University, Teachers College Columbia University, University of California, Berkeley, University of California, Los Angeles, University of Colorado, University of Michigan, University of Pennsylvania, and University of Wisconsin.

By Lorrie A. Shepard, William R. Penuel, and Kristen L. Davidson

When Congress passed the Every Student Succeeds Act (ESSA) in December 2015, it carried forward many of the same testing requirements that existed under No Child Left Behind (NCLB). But at the same time, it softened the consequences, taking away the federal government's power to determine what will happen to schools that fail to meet specific testing goals. From now on, states and districts can decide for themselves what achievement targets to set, and they can choose to focus on needed supports instead of sanctions for their lowest-performing schools.

Free from strict Adequate Yearly Progress accounting, many state and district education leaders are exploring new ways to collect information about student learning as well as new ways to use that information. Where NCLB mandated testing strictly for summative purposes (i.e., to judge how well students, teachers, and schools have performed), ESSA permits and even provides some funding to encourage states to develop balanced assessment systems. Such systems encompass not just summative tests but also local, formative assessments, which could include curriculum-embedded assessments designed to provide teachers with insights about instructional supports that are needed.

LORRIE A. SHEPARD (lorrie.shepard@colorado.edu) is distinguished professor of education, **WILLIAM R. PENUEL** (@bpenuel) is professor of learning sciences and human development, and **KRISTEN L. DAVIDSON** is a postdoctoral research associate, all at the University of Colorado Boulder, School of Education.

The question is: What guiding principles would help ensure the quality of these new, balanced assessment systems? Drawing on lessons learned over three decades of research and reform, we argue that state and local leaders should take the lead in designing new assessments guided by two core principles: First, make assessments coherent, integrating them with rich curriculum and effective instruction; second, ground this integration of curriculum, instruction, and embedded assessments in equity-focused research on learning.

Building coherent assessment systems

The idea of building a coherent system of assessments “from classroom to state” was first advanced in a National Research Council committee report, *Knowing What Students Know* (Pellegrino, Chudowsky, & Glaser, 2001, p. 9), which synthesized findings from contemporary research on both learning and educational measurement. Whether an assessment is meant to be used in a classroom or for state accountability, it should assess what is truly valuable for students to learn, such as core ideas and key skills from the various content areas (p. 248). By contrast, many classroom worksheets and multiple-choice tests that mimic state exams have reflected a negative kind of coherence, requiring students to answer superficial questions or recall simple facts.

Further, the report explained that assessments should be coherent not only *vertically* (i.e., the same standards and learning goals drive assessments at both the classroom and state accountability levels) but also *horizontally*. Horizontal coherence, at each level of the system, refers to the conceptual integration of assessments with a shared model of learning. At the state level, this means that accountability assessments must fully embody learning goals envisioned by standards. At the district level, assessments must be coherent with standards, curricula, and professional development. And at the classroom level, horizontal coherence requires that assessments be so thoroughly integrated with curriculum and instruction that the insights they provide can immediately be put to use. Thus, classroom formative assessments must be built on much more fine-grained models of learning than state-level tests. In order to provide the kinds of specific feedback and instructional supports that students need at intermediate stages of development, teachers need research-based tools that are attuned to the very specific ways in which student understanding develops in each academic domain.

For example, let’s say that a state or district standard specifies that students must learn how to make and defend reasoned arguments about informational

texts. And let’s say that school system leaders introduce assessments that are designed to help teachers gauge the kinds of support their students need in order to reach this standard. We would judge this system to be *horizontally* coherent if students are also given classroom assignments that call upon them to make well-reasoned arguments and if teachers use practices (such as asking complex questions as prompts for classroom discussions) that have been found to help students strengthen their own logic and use of evidence.

If, in turn, classroom assignments, course grades, and external accountability tests routinely ask students to demonstrate such reasoning, then the assessment system would be vertically coherent as well. In short, all parts of the educational system, including curriculum, instruction, and assessments of all kinds, ought to be working toward the same goals, helping move students toward shared definitions of what they ought to know and be able to do.

Unfortunately, the testing mandates of the past two decades have only made things less coherent (or coherent but not meaningful, insofar as they have fostered a teaching-to-the-test approach, aiming toward narrow curricular goals). Even in the face of those mandates, many teachers have experimented with various kinds of formative assessment, from exit slips to self-assessments, multidraft writing projects, and others. But in most cases, they have done so without much support from their districts or guidance about which approaches are most strongly grounded in research. And while these teachers’ uses of formative assessment may have delivered a strong message about the kinds of learning they value, their students have likely received very different messages from the high-stakes standardized achievement tests they have had to take.

Given new flexibility under ESSA, districts and schools now have an opportunity to design and implement coherent systems, with formative assessment having a more prominent role. To do so effectively, however, educators will need some background knowledge about the research that supports the main approaches to formative assessment. Thus, in the next section, we offer a condensed summary of four key assessment models (Penuel & Shepard, 2016), with a focus on their underlying theories of learning.

Grounding assessments in a model of learning

In order to be *coherent*, an assessment system should be based on a shared model of learning. It won’t be *effective*, though, unless that underlying model of learning has a valid basis. And it won’t be *equitable* unless it includes curricular supports for students and

All parts of the educational system ought to be working toward the same goals, helping move students toward shared definitions of what they ought to know and be able to do.



adequate preparation of teachers to help students meet learning goals. Its goals must be compelling, but those goals must also be reachable, and they must feature teaching practices that are consistent with what is known about student motivation, identity formation, and cognitive development.

In an early and widely influential review of the research on formative assessment, Black and William (1998) identified a number of distinct lines of scholarship in this area. But while they noted that differing approaches to formative assessment relied on very different learning models, they did not offer a way to integrate those perspectives. Nor did they show how disparate ideas about motivation, self-assessment, mastery, the giving of feedback to students, and other issues could be integrated into a coherent whole.

Since that time, many researchers have described formative assessment as though it were a single, coherent practice, without recognizing that the label refers to varied learning goals and theories that are not necessarily compatible with each other because they draw upon very different conceptual models. Here, we call out those differences, describing four distinct approaches that have been promoted as “formative assessment.” We argue that the latter two perspectives hold the greatest promise for supporting more ambitious and equitable, next-generation visions of teaching and learning, but we also point out that each of these approaches has limitations.

#1. Data-driven decision making

Data-driven decision making is most accurately portrayed as a policy *theory of action*. It relies on no specific model of learning but, rather, draws its inspiration from theories of organizational change (Deming, 1986; Senge, 1990). The idea is that educators should set specific learning goals, use interim or benchmark assessments (sometimes marketed as “formative assessments”) to check student progress toward reaching them, find new teaching strategies to address areas of weakness, and continue to monitor student progress over time.

Data-driven decision making assumes that teachers will know how to help students — or will seek training that shows them what to do — if the interim tests reveal that students are struggling. But this assumption has never been supported by empirical research findings. And researchers have found this approach to be especially ineffective in low-performing schools that tend to lack the capacity to adapt in this way (Elmore, 2003).

To date, most of the research on data-driven decision making has focused on the work of data teams (groups of educators tasked with analyzing test results). Findings show that, at best, such teams are able to identify which students are the most in need of help and which objectives are most in need of reteaching (Shepard, Davidson, & Bowman, 2011). However, because interim assessments offer little to no insight into the reasons why students are underperforming or how to help them, their use hasn’t been found to lead to improvements in teaching or learning.

Further, data-driven decision making sometimes goes hand-in-hand with the use of extrinsic rewards and punishments to pressure students to improve. For example, teachers or administrators might post the results of interim tests in the hallway, letting everybody know who’s on track and who still needs to get, say, three more items right to reach proficiency. Yet, research has largely discredited this approach to motivating young people to learn. When students are struggling, being told how far behind they are doesn’t help them move ahead. Moreover, the response to identifying students in need of more support has often been to create “pull-out” programs for such students, rather than promoting more equitable teaching in the regular classroom. Instead, students need meaningful opportunities to engage with the material, ask questions, try ideas, and receive useful guidance and feedback from teachers and peers.

#2. Strategy-focused formative assessment

Strategy-focused practices include various tools and techniques for engaging students in analyzing and improving their own work, such as ways to pose questions that invite classroom discussion about on-

Join the conversation

facebook.com/pdkintl
[@pdkintl](https://twitter.com/pdkintl)

What guiding principles would help ensure the quality of these new, balanced assessment systems?



going projects, guidelines for assigning students to revise papers, and rubrics for self- and peer-assessment.

Evidence suggests that when teachers have meaningful opportunities to learn and try such techniques, they can become more skilled at creating classroom environments in which students assume an active role in their own learning. One well-known example of this approach, the King's-Medway-Oxfordshire Formative Assessment Project — in which students were taught to assess and build upon their ideas, identify their own sources of intrinsic motivation, and monitor and regulate their own learning — was found to have significant and positive effects on students' engagement and academic progress (Black et al., 2003).

It's important to note, though, that strategy-focused approaches are not grounded in any particular theory of learning. Rather, they amount to a loose collection of all-purpose strategies by which teachers and students can assess their ongoing work. While those strategies could be used to help young people acquire a deep understanding of sophisticated academic content, they can just as easily be used to promote rote mastery of a shallow curriculum. When it comes to the goals of learning, the approach is agnostic and, for that reason, quite limited.

A much more effective way to practice formative assessment, we would argue, is to choose tools and measures that are connected to the specific field and its goals and purposes. As the next two strategies demonstrate, the most powerful assessment tasks engage students in the genuine practices of the given

content area and help them reflect on and understand what it means to become truly proficient in that area.

#3. *Sociocognitive formative assessment*

Sociocognitive approaches are meant to assess students' understandings and skills as they participate in increasingly sophisticated practices common to disciplinary experts. Further, because thinking and learning are presumed to be fundamentally social activities, assessment is grounded in "local instructional theories" of learning, whereby a sequence of instructional activities is devised to support the particular group of students in developing proficiency (Gravemeijer, 2004).

Instructional sequences are typically based on either a "learning progressions" (or "trajectories") approach, which aims to help students move toward specific disciplinary goals (Simon, 1995; Smith et al., 2006), or a "knowledge-in-pieces" (or "facets") view, in which learning is seen as a less orderly process, and more attention is paid to the specific problems students are trying to solve that require use of disciplinary knowledge (diSessa, 1988). In both cases, assessment materials are designed for the particular content area, with attention to the challenges that students typically face when studying the given material, as well as common approaches to helping them move forward.

In addition to gauging students' progress in mastering content knowledge, sociocognitive strategies aim to help them take on the dispositions and identities of the given field. Thus, this approach tends to favor assessment practices (such as collaborative inquiry, expertly facilitated questioning and discussion, and qualitative feedback) that allow teachers to pay attention to how students are (and are not yet) acting, thinking, and reasoning in disciplinary ways.

One example that has been found to be effective is the Inquiry Project, a three-year sequence of instructional units designed to progressively build upper-elementary students' understandings about the nature of matter (Smith et al., 2006). Another example, designed for the middle grades, is the Contingent Pedagogies project (Penuel et al., 2017), which gives teachers specific questions with which to elicit students' ideas about the physical world, as well as discussion prompts meant to get students talking about ideas and methods that are central to the study of Earth science.

Two key strengths of the sociocognitive approach to assessment are its discipline-specific learning goals and its well-articulated learning theory (Penuel & Shepard, 2016). Rather than telling students how many correct and incorrect answers they got on a test, the point is to reveal how they think about and try to solve specific problems that have been chosen precisely because they relate to key concepts in the

given subject area. In turn, this gives teachers useful insights into what students already know, specific ideas that confuse them, concepts they'll need to learn right away, and other things to consider as they decide how best to teach the material.

However, developing these kinds of fine-grained, subject-specific assessment tools requires a lot of expertise and resources. As a result, they are not yet available to many schools and districts, or for certain topics and grade levels. Further, since these tools are meant only to assess students' *understanding* of particular subject matter — and not to provide insight into their differing *values*, *experiences*, and *personal goals* — it remains to be seen whether their use will benefit all students equitably. It could very well be the case that such tools are indifferent to students' racial, ethnic, or gender identities, emerging bilingualism, conditions of poverty, and other important dimensions of students' lives.

#4. Sociocultural formative assessment

Sociocultural interventions share with the sociocognitive approach many of the same research-based premises about the social nature of learning and development, as well as a focus on student participation in disciplinary ways of knowing and doing. The two theories of learning diverge, however, in terms of how they account for students' diversity. In short, sociocultural approaches more explicitly allow for students to engage with academic content and practices through differing entry points and to follow differing pathways to mastery.

Sociocultural theories of learning recognize that

Many state and district education leaders are exploring new ways to collect information about student learning, as well as new ways to use that information.



when students arrive at school, they bring with them important knowledge and interests that should inform curriculum and instruction. Rather than curricula or instructional practices that ignore students' experiences, teachers ought to help students reflect on how the school's ways of knowing, doing, and being relate to the practices that are valued in their own families and communities (Bang & Medin, 2010). A key purpose for assessment, then, is to elicit information about students' experiences and help them relate their own interests and goals for learning to becoming a part of a disciplinary community.

One promising example is the Bellevue-University of Washington Curriculum Redesign Partnership, which takes units of study from the district's elementary science curriculum and repurposes them, altering them in ways that give students more agency in the classroom and that tap into their diverse interests in the given topics. For instance, some units have been redesigned to build on students' knowledge about their communities, such as by using photography to document the everyday lives of the people who live there (Clark-Ibañez, 2004). At the beginning of a unit on microbes and health, for example, students take photos of activities they do in daily life to prevent disease and stay healthy. Then they share these photos in class as a way to bring personally relevant experiences into the classroom as they launch the unit. Their documentation also helps shape a student-led investigation focused on students' own questions, which are refined as students encounter key ideas in microbiology.

Using students' interests, experiences, and knowledge is an important strategy in equitable instruction. In order to become an integral part of balanced systems of assessment, however, sociocultural interventions like the one implemented in Bellevue will need to be embraced by a wide range of stakeholders. To this end, educators and researchers need to communicate with families and other educational stakeholders about the value of this approach to teaching and how its outcomes can be mapped onto familiar disciplinary standards and learning goals.

Conclusion: Start with local curriculum and instruction

Under ESSA, states and districts have the opportunity to build coherent systems in which formative assessments are codeveloped and integrated with local curricula, instruction, and professional learning — all of which are grounded in the same research-based model of learning.

It will also be critical to design for coherence between local and state-level models of learning when possible or to acknowledge when a shared model of learning is not possible because of limitations in the

Join the conversation

facebook.com/pdkintl
[@pdkintl](https://twitter.com/pdkintl)

The most powerful assessment tasks engage students in the genuine practices of the given content area.



model implied by state tests. Painful lessons from the past remind us that creating a coherent and effective assessment system between classroom and state-house does not mean building a *single* instrument to serve both formative and summative purposes. In the 1990s, reformers seized upon the idea of creating large-scale “tests worth teaching to,” which were meant to drive instructional improvements even as they collected accountability data. Suffice it to say that this did not work as planned. To be valid, reliable, and affordable, state accountability tests must be standardized, and this often makes them ill-suited to serve as models for high-quality teaching and learning at the local level.

A wiser approach, we believe, would be to start *not* with statewide accountability tests as the primary driver of educational reform but to begin, instead, with local decisions about curriculum and instructional practices, informed by small-scale (and usually low-stakes) assessments that are grounded in a single, coherent model of learning that is consistent with contemporary research findings about cognition, child development, motivation, identity formation, and equity-focused instruction.

References

Bang, M. & Medin, D. (2010). Cultural processes in science education: Supporting the navigation of multiple epistemologies. *Science Education*, 94 (6), 1008-1026.

Black, P., Harrison, C., Lee, C., Marshall, B., & William, D.

(2003). *Assessment for learning: Putting it into practice*. Buckingham, UK: Open University Press.

Black, P. & William, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5 (1), 7-74.

Clark-Ibañez, M. (2004). Framing the social world with photo-elicitation interviews. *The Behavioral Scientist*, 47 (12), 1507-1527.

Deming, W.E. (1986). *Out of the crisis*. Cambridge, MA: MIT Press.

diSessa, A.A. (1988). Knowledge in pieces. In G. Forman & P. Pufall (Eds.), *Constructivism in the computer age* (Vol. 49-70). Hillsdale, NJ: Erlbaum.

Elmore, R.F. (2003). Accountability and capacity. In M. Carnoy, R.F. Elmore, & L.S. Siskin (Eds.), *The new accountability: High schools and high-stakes testing* (pp. 195-209). New York, NY: Routledge Falmer.

Gravemeijer, K. (2004). Local instruction theories as means of support for teachers in reform mathematics education. *Mathematical Thinking and Learning*, 6 (2), 105-128.

Pellegrino, J.W., Wilson, M.R., Koenig, J.A., & Beatty, A.S., (Eds.) (2014). *Developing assessments for the next generation science standards*. Washington, DC: National Academies Press.

Penuel, W.R., DeBarger, A.H., Boscardin, C.K., Moorthy, S., Beauvineau, Y., Kennedy, C., & Allison, K. (2017). Investigating science curriculum adaptation as a strategy to improve teaching and learning. *Science Education*, 101 (1), 66-98.

Penuel, W.R. & Shepard, L.A. (2016). Assessment and teaching. In D.H. Gitomer & C.A. Bell (Eds.), *Handbook of research on teaching* (pp. 787-850). Washington, DC: AERA.

Senge, P. (1990). *The fifth discipline: The art and practice of the learning organization*. New York, NY: Doubleday.

Shepard, L.A., Davidson, K.L., & Bowman, R. (2011). How middle school mathematics teachers use interim and benchmark assessment data. *CSE Technical Report 807*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Simon, M.A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, 26 (2), 114-145.

Smith, C.L., Wiser, M., Anderson, C.W., & Krajcik, J. (2006). Implications of research on children's learning for standards and assessment: A proposed learning progression for matter and the atomic-molecular theory. *Measurement: Interdisciplinary Research & Perspective*, 4 (1&2), 1-98.