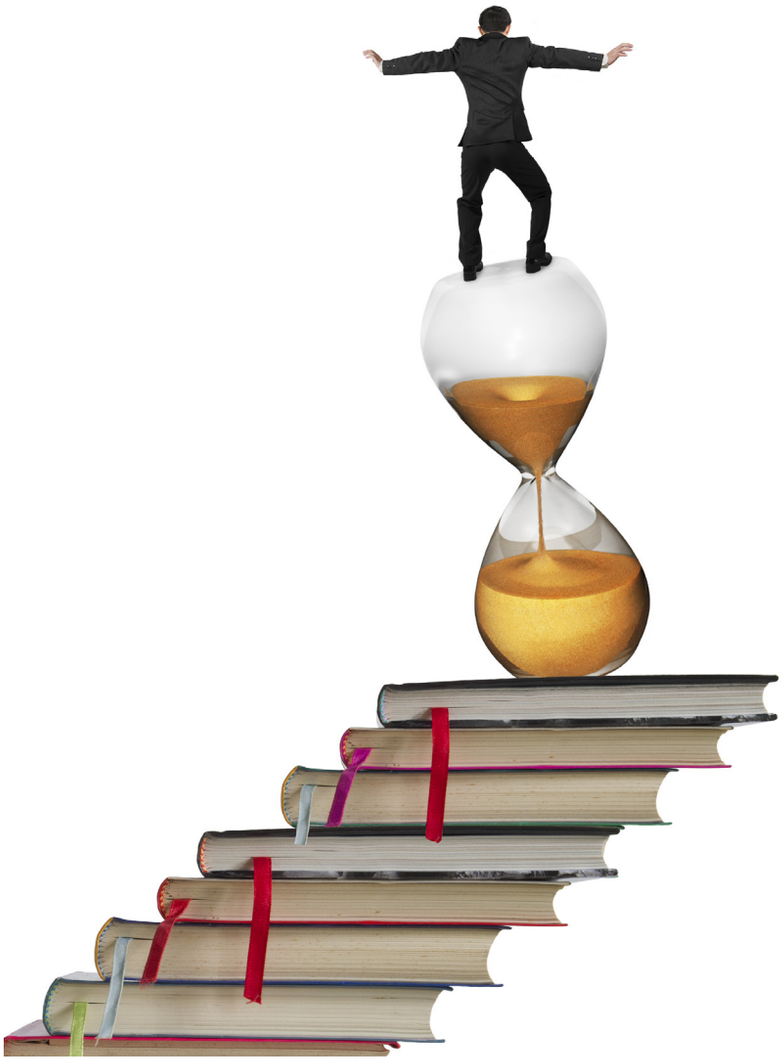


Not as Easy as It Sounds:
Designing a Balanced Assessment System

It's one thing to agree that no one test can tell us what we need to know about student learning. It's another to build a coherent system of assessments that has many parts and purposes.

by Rajendra Chattergoon and Scott Marion

Many states and school districts are rethinking how they do educational assessment. A few are going further: attempting to build "balanced," "comprehensive," or "next-generation" assessment systems. At the same time, practitioners and researchers have long mulled the purposes and parts such systems should have. But just a few have thought deeply about what a system of assessment actually means. In the words of educational psychologist Theodore Coladarci: "A collection of assessments does not entail a system any more than a pile of bricks constitutes a house."[1] This article focuses on building the house.

For states that are building, evaluating, or modifying their assessment systems, we offer three criteria to look for: coherence, a theory of action, and efficiency. We developed these criteria after reviewing literature on educational assessment systems, and we have been applying them in work with state and district leaders.

Coherence

There are two aspects to consider when evaluating coherence within an assessment system.[2] First, the assessments in a system must be compatible with the models of how students learn content and skills over time. Second, curriculum, instruction, and assessment must be aligned to ensure that the entire system is working toward a common set of learning goals.

Underpinning both of these features of coherence is a well-articulated theory of how students learn. Several researchers have written about these two dimensions of coherence.[3] However, most have only examined coherence among a relatively small group of classroom assessments and not within a state assessment system.

Models of student learning describe how knowledge and skills progress as students are exposed to instruction and other learning opportunities. They build upon content standards by describing a sequence of milestones that students should reach as they progress toward a learning target. Some states have started to adopt documents that describe progressions of student understanding of a topic across grades.[4] The development and validation of these learning progressions often require rigorous research into student cognition and assessment. As such, it may be challenging for states to draft their own models of student learning from scratch. It might be more promising for states to collaborate with one another and with researchers to identify promising models that may then be adopted statewide. Evaluations of state assessments should identify and evaluate the appropriateness of the models of student learning that a state endorses.

A major advantage of having well-articulated models of student learning is that such models can inform teachers' formative assessment practices.[5] And when curriculum, instruction, and assessment align in this way, the distinction between an assessment system and a classroom-level instructional system evaporates. One way to get inside the "black box" is to adopt or develop classroom observation protocols that are focused on holistically evaluating this alignment.

One example of such a protocol is the Danielson Framework.[6] A teacher's curriculum and its connection to models of student learning are evaluated as part of a "planning and preparation" domain. The domains of "classroom environment" and "instruction" address the quality of a teacher's interactions with students in a classroom. Although there is no separate domain for evaluating teachers' assessment practices or assessment literacy, some of the actions ascribed to strong formative assessment practices are evaluated in subsections. When coupled with an evaluation of the appropriateness of the models of student learning that a state uses, the results from teacher observations can provide a gauge of the system's coherence.

Both aspects of coherence are grounded in a theory of how students learn. While the research community has over the years engaged in lively debate about the merits of different theories, there is emerging consensus that learning happens through cognitive mechanisms within a social context.[7] This consensus matters, because different models of student learning point to different assessment tasks. A balanced assessment system should include a range of cognitively complex tasks. Consider, for example, a state assessment system in which a majority of tasks only require recall of basic facts. Such an assessment system cannot tap into students' abilities to learn information in collaborative or social contexts.

Theory of Action

The second criterion for a high-quality assessment system is that a well-articulated theory of action guides interaction among its component parts.[8] Such a theory of action includes the following:

- identification of an overall purpose for the system;
- identification of the stakeholders and what they need from assessments;
- a map of how assessments are being used to meet stakeholder needs and are valid for intended purposes;
- specification of the mechanisms (or processes) by which the assessments are designed to work and what needs to happen in order for the purposes to be realized; and
- evaluations of the extent to which stakeholders' needs are met.

A set of assessments, even if they cohere, will not fulfill the intended purposes if the information never reaches the intended user. Multiple sources of assessment information should be integrated as part of the assessment system design—not after the fact.

By developing a theory of action for its assessment system, a state defines its overall purpose and a plan for achieving that purpose.[9] We work with a state that is piloting a new accountability system that incorporates local and common

performance assessments. State policymakers shifted to performance assessments in order to document student learning of key concepts and skills and to bolster local instructional and assessment capacity. This theory of action intentionally minimizes the state's role in the accountability system and instead redirects resources to develop local capacity. In order to evaluate the effectiveness of this new system, the state is specifying how intended inputs and mechanisms produce expected outputs.

To evaluate the broader assessment system of which the accountability system is a part, the state would benefit from specifying how information should flow to the intended users. Figure 1 gives a hypothetical example. At the center of the diagram are elements that contribute to coherence. The diagram describes how information flows to four stakeholders, one of which is the state education leader. In this example, the state education leader requires information for accountability about the overall, aggregated performance of schools in the state. This information can be collected via standardized tests, indicators of students' opportunity to learn, and aggregated information about district "common assessments." To make decisions, state education leaders do not need the fine-grained information teachers do about student performance on the common assessments or information about students' daily performance on classroom tasks. This diagram illustrates one way an assessment system can meet the needs of those who use assessment results.

Attention to efficiency enables each assessment to do what it is designed to do.

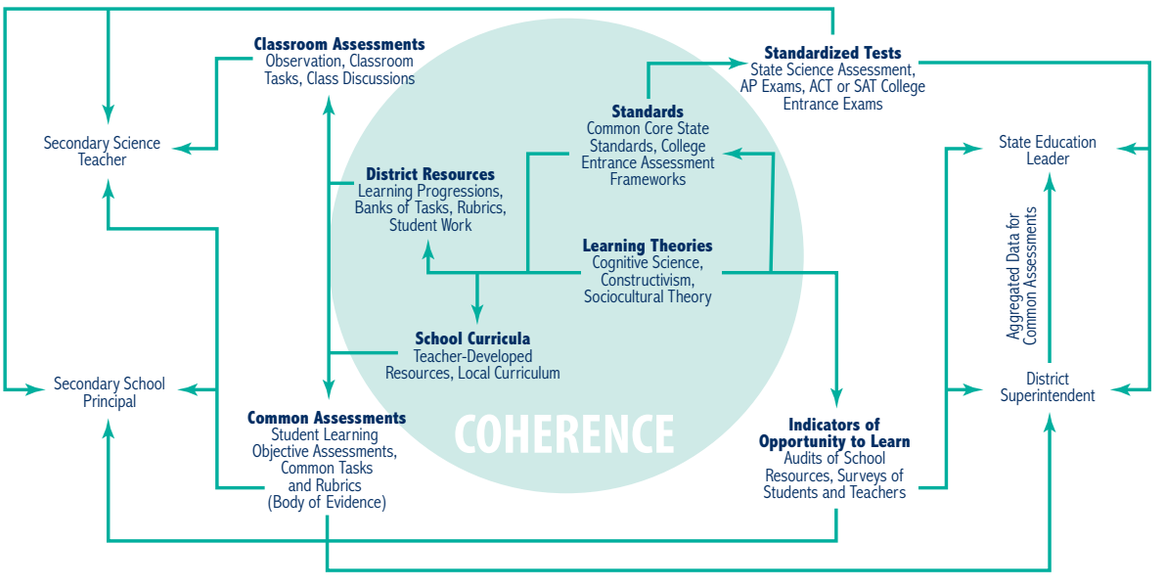
Efficiency

An assessment system has to provide users with a clear picture of what students know and are able to do. It can't do so if parts of the system are not working together efficiently. Assessment efficiency means getting the most out of assessment resources and eliminating redundant, unused, and untimely assessments. Attention to efficiency enables each assessment to do what it is designed to do. For example, if a state wants to give educators information to help them adjust instruction, its assessments must be tied to the curriculum that is being used. These assessments should in turn yield timely, detailed information about the knowledge and skills being assessed at the local level.

Another partner of ours, a large school district in the southern United States, is seeking to streamline its assessments. Teachers in the district must administer more than a dozen assessments throughout the academic year, and many of them multiple times each year. For example, during October, high schools must administer a college entrance exam (e.g., the PSAT, SAT, or ACT) and an interim assessment, alongside any local assessments that teachers or teams of teachers develop. Recognizing that this structure was unwieldy, the district has begun initial planning and evaluation to make the assessment program more efficient.

States need to focus on assessment efficiency to accomplish a couple of things:

Figure 1. Hypothetical Information Flow for an Assessment System



- Identify and reduce assessments that do not cohere with the state's models of student learning and that are not mandated.[10] Carefully defining the intended purposes and uses for each assessment in the system and what will be required for each assessment to be able to fulfill the intended purposes will be a prerequisite for this task.
- Define the minimum number of assessments necessary to meet the information needs of all stakeholders.

Although this district has not formally endorsed a specific model of student learning, the state has provided some limited resources to help districts think about learning progressions. In the absence of clear models, it is difficult for a district to decide which assessments to shed. However, one way to increase efficiency is to evaluate how the information generated from assessments is used. For example, how does the district use information from the October interim assessments? Can this information be obtained from other assessments within the system?

Advancing Assessment Reform

We have described three characteristics of good assessment systems to aid states in modernizing their own assessments. Even though we have described them discretely, successful reform movements will likely be attending to all three concurrently. Coherence is the kernel that should be present for any reform to succeed. Without coherence, it does not make sense to speak of pathways in which information flows or an efficient set of assessments.

A few words of caution follow. An assessment system inherently serves multiple purposes: accountability, evaluation, instructional improvement, and monitoring, for example. However, meeting all these purposes requires thoughtful planning about which data will be privileged at each level. For example, what types of assessment data will be used for accountability and which will be reserved for, say, evaluation? Many states and school districts use large-scale summative tests for accountability and other assessments for the other purposes. Given the need for comparability across districts in making state policy decisions, this allocation makes sense. However, serious unintended, negative consequences can ensue from layering too many requirements onto the state-mandated summative test. There is a long, unhappy history of accountability tests driving and narrowing the curriculum as educators focus their efforts on the tests that get counted.[11] More important, statewide summative tests may create an unnecessary divide between state and local assessments. As states revisit their assessment programs, they would do well to think about how changes will affect districts, schools, and classrooms and how coherent, balanced assessment systems could best be supported. ■

[1]T. Coladarci, "Is It a House…or a Pile of Bricks? Important Features of a Local Assessment System," *Phi Delta Kappan* 83, no. 10 (2002): 772–74.

[2]We draw on the definition of coherence in an assessment system provided by the National Research Council, *Knowing What Students Know: The Science and Design of Educational Assessment* (Washington, DC: National Academies Press, 2001).

[3]For a collection of studies on coherence in classroom assessment, see M. Wilson, *Towards Coherence between Classroom Assessment and Accountability. The One Hundred and Third Yearbook of the National Society for the Study of Education, Part II* (Chicago, IL: National Society for the Study of Education, 2004).

[4]For example, the Department of Public Instruction in North Carolina is working on providing learning progressions for Common Core English language arts (http://elaccss.ncdpi.wikispaces.net/) and math (http://maccss.ncdpi.wikispaces.net/Home).

[5]North Carolina's Office of Early Learning is using "construct progressions" as part of the professional development in formative assessment that they are providing to school districts. For details about this project, see http://r5k3formativeassessmentsupport.ncdpi.wikispaces.net/.

[6]For a description of Charlotte Danielson's Framework, see https://danielsongroup.org/framework/.

[7]For a thorough description of how learning theories connect to assessment, see L. Shepard, "The Role of Assessment in a Learning Culture," *Educational Researcher* 29, no. 7 (2000): 4–14.

[8]See R. E. Bennett, "Cognitively Based Assessment of, for, as Learning (CBAL): A Preliminary Theory of Action for Summative and Formative Assessment," *Measurement: Interdisciplinary Research and Perspectives* 8, no. 2-3 (2010): 70–91. Another good resource is Scott Marion's presentation on using theories of action for assessments in a policy context: http://www.nciea.org/publication_PDFs/RILS_SM2010.pdf.

[9]One resource for developing theories of action is Erika Hall's "Framework to Support the Validation of Educator Evaluation Systems": http://www.nciea.org/publication_PDFs/A%20Framework%20to%20Support%20the%20Validation%20of%20Educator%20Evaluation%20Systems_EH071114.pdf.

[10]Organizations are developing resources to help inventory state and district assessment systems. For example, Achieve has developed an "assessment inventory" for school districts: http://www.achieve.org/assessmentinventory.

[11]For more information on the history of accountability testing, see L. Shepard, "A Brief History of Accountability Testing, 1965-2007, in K. E. Ryan and L. A. Shepard, eds., *The Future of Test-Based Educational Accountability* (New York: Routledge, 2008).

Serious unintended, negative consequences can ensue from layering too many requirements onto the state-mandated summative test.

Rajendra Chattergoon is a doctoral student in the research and evaluation methodology program in the School of Education at the University of Colorado Boulder, and Scott Marion is executive director at the National Center for the Improvement of Educational Assessment.