

COHERENCE: KEY TO NEXT GENERATION ASSESSMENT SUCCESS

BY JOAN L. HERMAN



Assessment and Accountability
Comprehensive Center

AACC • A WestEd and CRESST partnership

AACC: Assessment and Accountability Comprehensive Center: A WestEd and CRESST partnership.
aacompcenter.org

Copyright © 2010 The Regents of the University of California

The work reported herein was supported by WestEd, grant number 4956 s05-093, as administered by the U.S. Department of Education. The findings and opinions expressed herein are those of the author(s) and do not necessarily reflect the positions or policies of AACC, WestEd, or the U.S. Department of Education.

To cite from this report, please use the following as your APA reference:

Herman, J. L. (2010). *Coherence: Key to Next Generation Assessment Success (AACC Report)*.
Los Angeles, CA: University of California.

The author thanks the following for reviewing this policy brief and providing feedback and recommendations:
Ron Dietel (CRESST); and for editorial and design support: Tamara Lau and Amy Otteson (CRESST).

COHERENCE: KEY TO NEXT GENERATION ASSESSMENT SUCCESS

JOAN L. HERMAN

Close your eyes and ask yourself what's wrong with current assessments.

Assessments don't...

- ...adequately measure complex thinking and problem solving,
- ...provide results fast enough to help inform instruction,
- ...give English learners, students with disabilities, or traditionally low performing students a fair chance to show what they know or how they've progressed;
- ...reward good teaching, but instead narrow curriculum and encourage "teaching to the test."

You could add your own testing concerns to this list. These problems and numerous others, I think, can be captured in two broad statements:

- Current tests don't measure the "right stuff" in the right ways, and
- They don't well serve the purposes we need or want them to serve.

What can we do differently?

In a single word but with many steps, I suggest the word "coherence." I believe that by making our assessments more coherent in both design and use, we can create assessment systems which will measure the *right stuff* in the *right ways* while better serving intended purposes, particularly the purpose of improving teaching and learning. The current Race to the Top Assessment Program (RTT) provides states a sizeable carrot—\$350 million—to do just this, creating a next generation assessment system that reflects new Common Core State Standards and supports accountability and improvement at all levels of the educational system: state, district, school, classroom.

The way forward to better assessment begins with the conception of assessment not as a single test but as a coherent *system* of measures. Coherent systems must be composed of valid measures of learning and be horizontally, developmentally, and vertically aligned to serve classroom, school, and district improvement.

WHY A “SYSTEM” OF ASSESSMENTS?

Research on educational testing provides ample evidence of the shortcomings of trying to rely on a single annual state test to serve accountability and improvement purposes (National Research Council [NRC], 2001). A sole multiple choice test, administered in an hour or two, cannot cover the full range of year-long standards representing what students should know and be able to do. Picture, for example, the typical test: it uses a collection of individual test questions, each addressing different, discrete aspects of learning. If the questions do not fully represent the standards—as research suggests (Resnick, 2006; Webb, 1997)—then the situation is like trying to understand an artist’s work by examining only a few, disconnected pieces of it, or by watching only the first act of a three-act play. The pieces you see may be important, but nonetheless miss essential elements of the whole (see Figure 1).

In contrast, a system composed of multiple assessments can illuminate a broader, deeper perspective of student knowledge and skills. A second assessment for example, cannot only assess more content knowledge, but, if designed to measure *applied* knowledge, can evaluate different types of skills.

Although it is an overused example, a driver’s license test illustrates a coherent, multi-assessment system. States typically use a written, multiple-choice test to measure our *rules-of-the-road* knowledge, such as recognizing signs at intersections, knowing how much space to leave between your car and the one in front of you, or at what distance to start signaling before making a turn. States use a performance test to measure our ability to *apply* the rules in a real situation, driving a car. Do we fully stop at a stop sign? Can we parallel park? Do we scan the road for possible hazards as we drive?

Knowing the rules of the road may be an essential prerequisite to being a good driver, but having that knowledge doesn’t ensure capability to apply it. My 93-year-old mother, for example, knows all the rules and can pass the written test, but the state wants to be sure that she can still apply that knowledge through a driving test.¹

So too with educational tests. An assessment system comprised of multiple types of measures can provide a more thorough picture of student learning. Such systems also can be more responsive to the diverse decision-making needs for those who need data to support improvement—teachers, administrators,

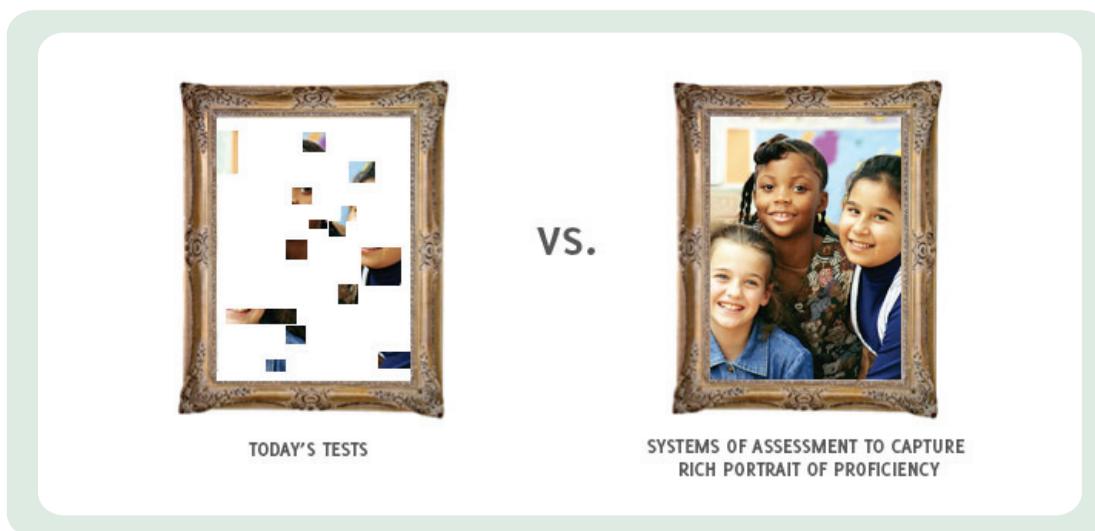


Figure 1. Seeing the full picture

¹ Our example is for illustrative purposes only. California requires drivers over 70 to retest if they are involved in two or more accidents in one year.

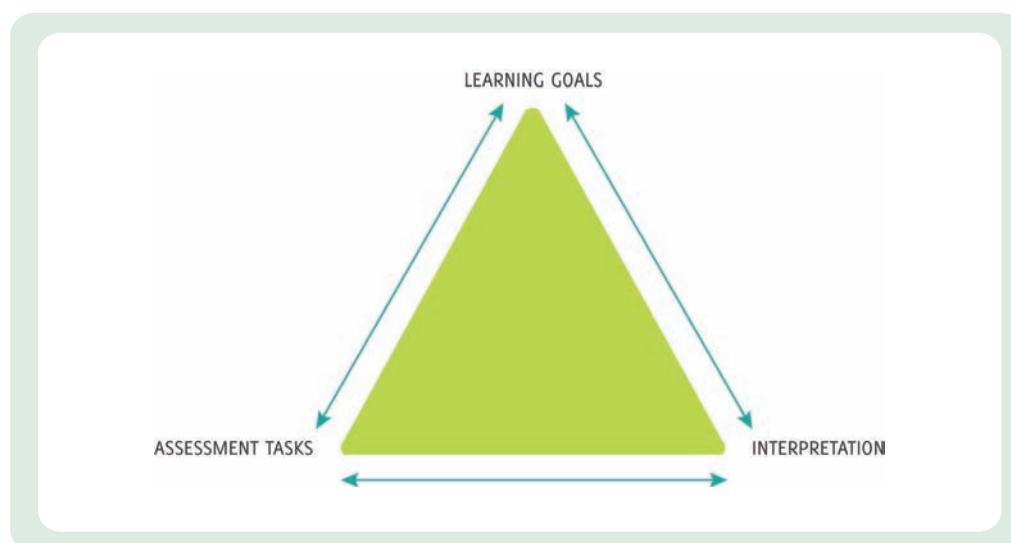


Figure 2. Learning and assessment model

parents, students. A solitary, end-of-year test simply cannot provide sufficient formative information to guide teaching and learning throughout the year.

FUNDAMENTAL COHERENCE WITH SIGNIFICANT LEARNING

Coherent assessment systems are comprised of component measures that each reflect significant learning goals and provide accurate information for intended purposes. Drawing from the *Knowing What Students Know* National Research Council conception (National Research Council [NRC], 2001), coherence starts with a clear specification of the *goal(s)* to be measured (see Figure 2). Next, *assessment tasks* are specially designed or selected to reflect the learning goal(s). Finally, an appropriate *interpretation* framework is applied to student responses to reach valid conclusions about student learning—for example, a score of “proficient” on a state test or an inference about the source of a student’s misunderstandings in teachers’ formative practice.

The quality of an assessment—termed validity by the measurement community—resides in part in the relationships among and between the three vertices. For example,

- Are the assessment tasks aligned with significant learning goals? Fair and free from bias? Accessible for all students?
- Does the interpretation of student responses to the task(s) yield accurate inferences about student learning? Does the interpretation support the intended purpose(s)?
- Does performance on the assessment reflect important capability? Does it transfer to other settings or applications beyond the assessment?

It is worth underscoring that assessment development starts with essential goals and creates assessment tasks and items to specifically reflect those goals—and not vice versa.

Moreover, it is important to remember that beyond providing data to inform decision-making, assessments also signal to teachers and students what is important to teach and learn, plus what kinds of knowledge are valued. In light of this signaling function, it is important to ask:

- Are the assessments worth teaching to?
- Do they model and communicate meaningful teaching and learning?

HORIZONTAL COHERENCE

Horizontal coherence involves the close alignment of learning goals, instruction, and assessment (see Figure 3), an essential synchronicity in the use of assessment to improve learning. Teachers start with specific learning goals, engage students in instructional activities to reach those objectives, and use assessment to get ongoing feedback on how students are doing. Teachers and students then use the feedback to close the gap between where students are and where they are expected to be. Similarly, teachers, schools, or districts may use assessment data periodically to take stock of how students are performing, analyze curriculum strengths and weaknesses, identify promising practices and those who may be struggling, then use this feedback to strengthen programs, teaching, and learning.

DEVELOPMENTAL COHERENCE

Complementing horizontal coherence, developmental coherence is the extent to which learning goals, instruction, and assessment are continually intertwined *over time* to promote student progress. Because different types of assessments may be given during various times of the year, developmental coherence also involves the extent to which these assessments are coordinated to support the same, significant goals. Developmental coherence means that daily goals build to weekly and unit learning objectives. These, in turn, lead to important quarterly accomplishments, then to yearly grade level standards, and finally, over many years, to college and work readiness.

Assessments serving various users support this same progression: teachers' on-going formative assessment processes on a daily and weekly basis provide feedback that supports student learning toward school

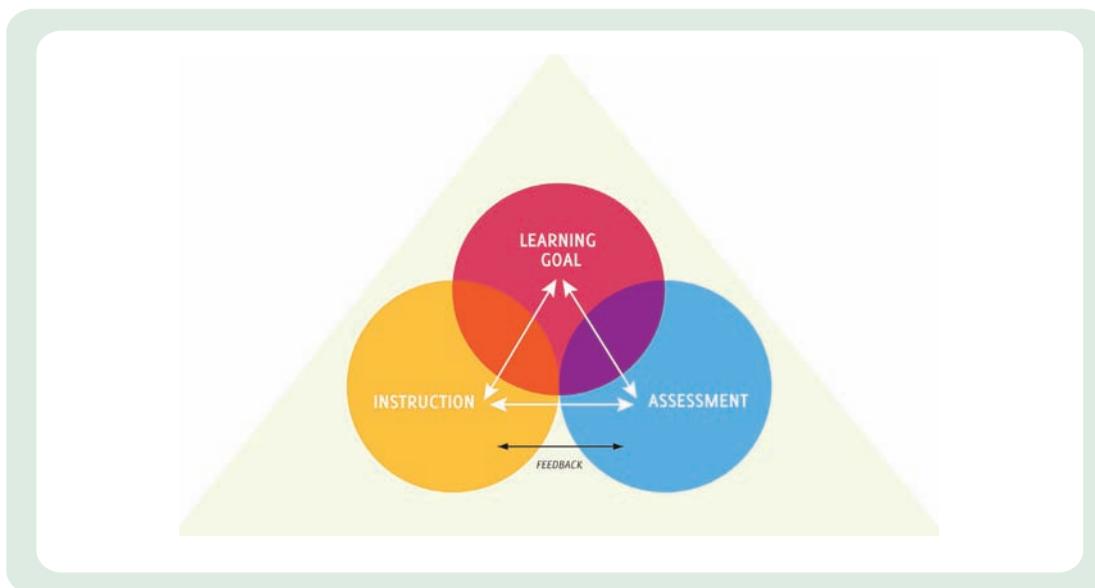


Figure 3. Horizontal coherence

A horizontally coherent assessment system can detect good teaching and improved learning if the assessments are sensitive to instruction: If students have been taught effectively and have learned the requisite content and skills, the learning should be evidenced in higher test scores. While sensitivity to instruction seems obvious, it cannot be assumed.

benchmark assessments. Feedback enables educators to refine their efforts toward end-of-year standards and annual accountability tests. Today builds to tomorrow, tomorrow builds to the next day, and onward (see Figure 4), with the important proviso that all of these assessments are fundamentally coherent with important learning (see Figure 2).

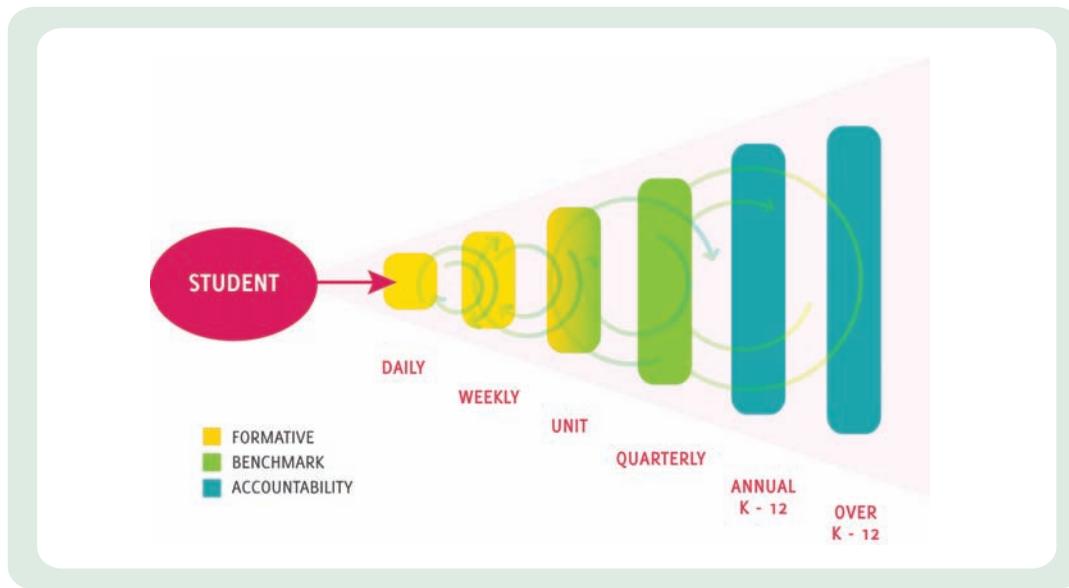


Figure 4. Developmental coherence

VERTICAL COHERENCE

Figure 5 builds on our assessment system model in two important ways. First, it highlights that the system must serve decision-makers at multiple levels—classroom, school, district, and state—and secondly, it introduces *use* as a critical model component. For assessment to support learning, results must not only be learning-based and provide relevant information for decision-makers, but must actually be *used* to make changes that will improve achievement.

Classroom teaching and learning, as Figure 5 demonstrates, is the ultimate target for assessment that supports improvement. It is also the place at which assessment is most frequent—ideally, as part of teachers’ ongoing, formative practice of continuous improvement. School or district assessments are more periodic, with feedback being used to support decision-making by teachers, schools, and districts. For example, teachers jointly analyze student work and determine implications for next steps; school or district administrators use results to identify needs for professional development, curriculum revision, and special interventions for some students

or teachers. State level testing usually provides an annual accounting of how students are doing, with implications for the distribution of rewards and sanctions, identification of general strengths and weaknesses in curriculum, program evaluation, and so on.

A central point in Figure 5 is that assessments at each level emanate from the same set of goals, communicate a shared vision of what is important for students to know and be able to do, and push teaching and learning in a common direction. The combination of assessments provides mutually complementary views of student learning that together reinforce important goals while strengthening coherence and validity of the entire system.

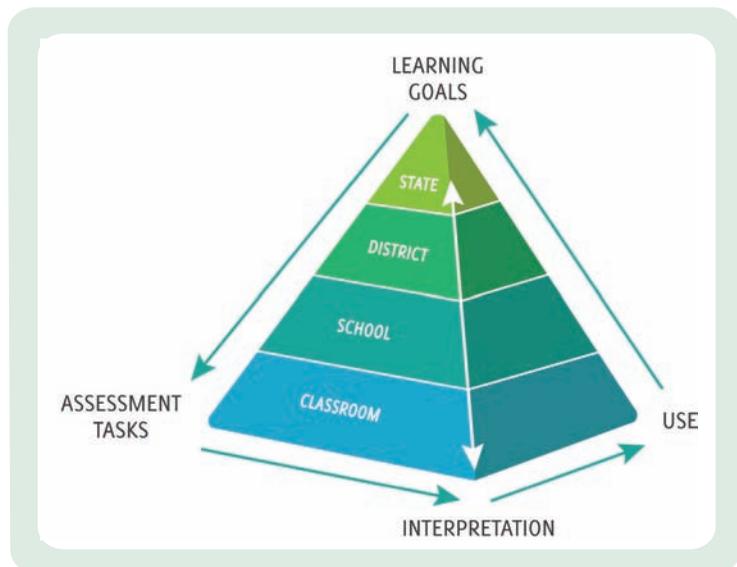


Figure 5. Vertical coherence

MODEL APPLIED TO RTT EXPECTATIONS

Federal expectations for the RTT assessment consortia lay out an ambitious set of purposes for next generation state assessment systems (see Table 1). The RTT emphasis on accountability testing turns our assessment system model on its head, leading with annual testing at the top, while classroom assessment is at the bottom. While this leaves a more fragile base for classroom teaching and learning, the emphasis on a system of assessments by the addition of through-

course exams to complement end-of-year assessments is very promising. Through-course exams—more extended, performance-oriented assessments conducted during the course of instruction—provide rich opportunities to assess students’ thinking and reasoning as well as their ability to apply and communicate their knowledge and skills in solving complex problems. Performance assessments also provide useful models of effective teaching while supporting authentic instruction and student learning.

ASSESSMENT	ASSESSMENT TYPE	PRIMARY USERS	USE - BASED ON RTT
Annual	On-demand annual	<ul style="list-style-type: none"> • State • District • Schools • Teachers • Parents • Students • Public 	<ul style="list-style-type: none"> • Teacher/Principal/School effectiveness • Professional Development Needs • School and District Quality • General feedback, both curriculum and student strengths/weaknesses • Recognize and build on excellence • Status/growth toward college readiness
Through-Course Exams	End of Unit Mid-Term Semester End-of-Course	<ul style="list-style-type: none"> • Schools • Teachers • Students 	<ul style="list-style-type: none"> • Assign grades • Inform short and medium term decisions about curriculum and instruction • Identify struggling students
School/District	Benchmark	<ul style="list-style-type: none"> • Districts • Schools • Teachers • Students 	<ul style="list-style-type: none"> • Inform short and medium term decisions about curriculum and instruction • Identify struggling students • Identify struggling teachers • Identify struggling schools • Identify promising practices • Identify year-to-year trends
Classroom	Formative Curriculum-embedded Student Work Discourse Discussion	<ul style="list-style-type: none"> • Teachers • Students 	<ul style="list-style-type: none"> • Inform immediate and short-term teaching and learning • Identify struggling students

Table 1². Assessment Purposes

² Table 1 was created based on a review of the expectations in the Race to the Top Assessment Program (2010). See Comprehensive Assessment System grant, <http://www2.ed.gov/programs/racetothetop-assessment/index.html>

MANY ASSESSMENT PURPOSES

Some of the purposes shown in Table 1 may be in conflict. For example, the need for reliable measures of student growth for teacher evaluation and the need to gauge student growth toward college readiness narrow the breadth and depth of learning that can be assessed, in that current growth methodologies require comparable content across measures. Algebra and geometry, for example, are two very different subjects, as are biology and physics. You can't get a good measure of students' growth in science by comparing their performance in biology one year to their performance in physics the next year. At the same time, narrow assessments that are consistent in content from year to year may not be worthy targets for classroom teaching and learning nor adequately represent progress toward college readiness. Teacher evaluation schemes that put teachers in competition may work counter to building professional learning communities that can best support teachers' capacity to improve student learning.

The National Research Council observed: "...the more purposes a single assessment aims to serve, the more each purpose is compromised...assessment designers and users [need to] recognize the compromises and trade-offs..." (National Research Council [NRC], 2001, p. 53). The same is likely to be true of systems of assessment. To the extent that different components of the system address different, potentially conflicting purposes and emphasize different goals, system coherence may be lost. When the various components of the system push educators in different directions, stasis may be the result.

DESIGN CHALLENGES

Determination of quality and effectiveness involves any number of interrelated design and validation questions. For instance, to what extent does the system, its individual and collective measures:

- Signal significant learning goals? Or the full range of expected standards?
- Reflect a coherent view of learning and how it develops? Or of common expectations for learning?
- Provide accurate information for intended decision-making purposes?
- Enable all students to show what they know and to demonstrate progress?
- Show sensitivity to instruction?
- Support intended use? By intended users?
- Maximize positive consequences and minimize unintended, negative consequences? What are the consequences for individuals from special subgroups including English language learners and students with disabilities?

CONCLUSION

Similar to sending a manned spacecraft to Mars, simultaneously answering the preceding questions requires creative design and comprehensive engineering, moving beyond the current state of the art. Otherwise, resources will be wasted and our next generation assessment systems will fall short of our expectations for them.

Ultimately, our goal is not to create the most sophisticated assessment system in the world—though that could happen. Our objective is to create systems that support educational improvement, better education for all students, so that every student is prepared for college and success in life.

*And, one-day we can ask ourselves,
What's right about assessment?*

REFERENCES

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy of Sciences.

Resnick, L. B. (2006). Making accountability really count. *Educational Measurement: Issues and Practice*, 25(1), 33–37. Retrieved from http://www.lrdc.pitt.edu/publications/pdfs/200807161455540.2006_Making%20Accountability%20Count.pdf

Webb, N. L. (1997, January). *Determining alignment of expectations and assessments in mathematics and science education*. (NISE Brief Vol. 1, No. 2). Retrieved from <http://facstaff.wcer.wisc.edu/normw/1997alignment-brief.htm>