



Center for
K–12 Assessment
& Performance Management

An independent catalyst and resource for the improvement of measurement and data systems to enhance student achievement.

Exploratory Seminar:
Measurement Challenges Within
the Race to the Top Agenda
December 2009

RESPONSES TO SESSION 4: DESIGNING HIGH-QUALITY INSTRUCTIONAL GUIDANCE ASSESSMENT SYSTEMS

This policy brief is based on reactions by Randy Bennett (Educational Testing Service), Cheryl Krehbiel (District of Columbia Public Schools), and Ray Pecheone (Stanford University) to the Session 4 presentations at the Exploratory Seminar: Measurement Challenges Within the Race to the Top Agenda, December 2009. Download a copy of the final papers written by Dr. Bennett and Ms. Krehbeil, as well as the other papers presented at the seminar, at <http://www.k12center.org/publications.html>.

RANDY BENNETT, EDUCATIONAL TESTING SERVICE

The papers on assessment for instructional uses correspond to work at ETS on new approaches to K-12 assessment. The overarching principle for that work is that *summative, interim, and formative assessments should be designed as complementary components of the same coherent system*. One cannot talk meaningfully about instructional guidance assessment systems without paying attention to summative assessments because the other forms of assessment always gravitate toward the summative. Each of the presenters articulated this principle in one form or another. We have developed eight consequent principles:

- Summative, interim, and formative assessments should be built from the same strong conceptual base. The base should consist not only of curriculum, as in Lauren Resnick's formulation, but also of results from cognitive-scientific research. That cognitive-scientific research shows how domain components can work together and how learning progressions show changes over time in the domain components.
- Summative and interim assessments should be designed to support instruction to the maximum degree possible (without compromising their primary purposes). These assessments should be worthwhile educational experiences that students learn from so that *teaching to the test* focuses on the competencies represented by the test, not specific test content or the test format. These assessments also should model good instructional practices and point teachers and students to areas that need follow-up.
- Formative assessment should use a range of task types and modes.

Created by Educational Testing Service (ETS) to forward a larger social mission, the Center for K–12 Assessment & Performance Management has been given the directive to serve as an independent catalyst and resource for the improvement of measurement and data systems to enhance student achievement.

- Formative assessment should use technology when technology can make a meaningful contribution.
- Formative assessment should not be used for accountability purposes, a point that the presenters did not address. Attaching consequences to formative assessments can reduce the value of formative assessment for teaching and learning purposes.
- Teachers and students should have maximum flexibility in the use of formative assessment, which the presenters implied indirectly.
- By design, students should be encouraged to take part in formative assessment, as Margaret Heritage implies in her assessment framework.
- Summative, interim, and formative assessments should be development experiences for teachers. The assessments, and the cognitive domain models and learning progressions on which the assessments are based, should serve to deepen teachers' understanding of the domain. The presenters could have emphasized this principle more.

CHERYL KREHBIEL, DISTRICT OF COLUMBIA PUBLIC SCHOOLS

From the “weeds,” so to speak, the view of someone working daily in classrooms with teachers in a distressed urban school system is that the instructional assessment ideas in the papers would be wonderful to implement, but is there time to work them out?

Theory is not going to save the lives of the children we work with, and these theories must work under the time constraints of the Race to the Top funding. Once selections of recipients of the funds are made, the winners will have only two years to spend the money. Are these ideas ready to go? If not, what is the measurement community going to say to the decision makers at the U.S. Department of Education? The Race to the Top guidelines say that summative assessments must provide information that will be used to evaluate teacher performance. If the measurement community has a problem with that, the community had better figure out what to do now because the bus is leaving. We may make some mistakes, but there always are casualties in war, and our students' lives are at stake.

Because of the urgency, how can the measurement community inform the process wisely? We must continue to work on the grand schemes because research must inform our work as we move forward, but let's not sacrifice good for perfect.

RAY PECHEONE, STANFORD UNIVERSITY

Like the creators of the dinosaurs in the movie *Jurassic Park*, we have not been paying much attention to the ecosystem in which assessment policies operate. That theme dominated the presentations and discussions over the two days of the Exploratory Seminar, but in very different ways.

In Day 1, what one took away was the idea that all indicators of learning are fallible and that our conversations about them need to be more transparent. We need to tell more truth to power and communicate the limitations of the work we are doing as well as some of the opportunities. I did not

walk away from the day thinking our processes are fatally flawed, but I did think we had to put them in context.

Day 2, on the other hand, was a rebirth and uplifting. The presenters helped us focus on the ecosystem, the work around teaching and learning. This requires a different policy structure, and I have five perspectives on that:

- *An assessment system should be grounded in a thoughtful, standards-based curriculum.* This was said many times. Instead of trying to disentangle curriculum, assessment, and instruction, they need to be linked. The glue for this is learning progressions.
- *All stakeholders and, particularly, all teachers should be involved in the development of the standards and the assessments, not selective committees.* This can bring coherence in a school, an environment where everyone is pulling together.
- Assessments should include evidence of actual student performance on challenging tasks that go beyond what currently is being tested.
- Teachers need to be involved in the scoring process to build their knowledge and understanding of what challenging tasks look like in student work. The assessments should have curriculum embedded in them. As others have said, more curriculum-based assessments are a better choice than standards.
- *An assessment system should rely on multiple measures in multiple formats—and lower the flame on the nature of stakes.* Performance assessments should be tied to more moderate stakes and used in combination with other indicators of learning, such as grades, course exams, graduation requirements, and college admissions.

The federal government could call for the designs of the next generation of assessments to take multiple measures seriously and include all forms of assessment, as well as provide funding to build the capacity of teachers and administrators to do the work. I propose that it also support a *curriculum and assessment resources center*, which would be a depository of images of what high-quality instruction looks like. This different kind of research center would provide exemplars that can be used to design curriculum and that can be adapted and integrated into the curriculum as needed.

State policies ought to center on the integration of curriculum, instruction, and assessment, using the Common Core of Standards, and support districts in linking this integration to the outcomes they have decided upon. The policy environment requires that the system be worked from both ends and in all dimensions.

Major Discussion Points

While there was general agreement in favor of learning progressions, the participants in the discussion of the presentations seemed still unsure of the definition of learning progressions and their purposes. If they should frame the curriculum, then wouldn't the term *curriculum sequences* be more appropriate?

Learning progressions are an attempt to lay out long progressions in stages, explained one participant, “and we are still in a phase of clarifying our meaning and what grain size to use to explain these things.” The steps on a progression are really far apart, which means they are *coarse grain*. One comment was a caution: “We may have the same danger with learning progressions as with formative assessments—appropriating them for policy purposes when we don’t yet understand them.”

As in previous discussion, the capacity of teachers to play a role in developing assessments and dealing with standards was an issue. Can teachers produce good formative assessments on the fly? Can they take the laundry list of standards and build schema out of them? One reply was that schema theory needs to be put aside in favor of the current interest in stages of development in learning. If the standards are put into schema, the result will be the current multiple-choice tests.

There was an argument in favor of continuing with current summative assessment systems while new approaches linking curriculum, standards, and assessments are being developed. But, one discussant asked if the money is going to be in summative assessments or in curriculum development. Another pointed out that enough is known about fixing summative assessments, knowledge that can be shared across states. Small things can be done to mitigate some of the unwanted consequences of assessment systems, it was pointed out, such as making items less predictable and using tests that span across a broader domain.

The issue of urgency was considered important, but everything depends on having coherent standards, preferably tied to performance indicators. And, warned a presenter, “We need to make sure that what we have to do immediately does not contradict where we want to go.”

For More Information

For more information on this subject, please see the papers by Dr. Bennett, Dr. Heritage, Ms. Krehbiel, and Dr. Wilson:

Bennett, R. (2010). *Recommendations for high quality instructional guidance assessment systems and how they might articulate with an accountability system: Discussion*. Retrieved from <http://www.k12center.org/publications.html>.

Heritage, M. (2010). *Assessment for teaching and learning*. Retrieved from <http://www.k12center.org/publications.html>.

Krehbiel, C. (2010). *Comments on papers presented by Margaret Heritage and Lauren Resnick: Recommendations for high-quality instructional assessment systems*. Retrieved from <http://www.k12center.org/publications.html>.

Wilson, M. (2010). *Assessment for learning and for accountability*. Retrieved from <http://www.k12center.org/publications.html>.