



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

ASSESSMENT FOR LEARNING AND FOR ACCOUNTABILITY

This policy brief is based on a presentation by Mark Wilson (University of California, Berkeley) at the Exploratory Seminar: Measurement Challenges Within the Race to the Top Agenda, December 2009. Download a copy of the final paper written by Dr. Wilson, as well as the other papers presented at the seminar, at <http://www.k12center.org/publications.html>.

Assessment for learning and assessment for accountability should not be separated. The way to bring them together is to reverse the current strategy of moving from content standards to large-scale, or summative, assessments, and then to classroom-based, or formative, assessments. The flow should be from an improved set of standards to classroom assessments and then to large-scale assessments. This can be done by using learning progressions and learning performances.

Those who actually build assessments—state testing directors and testing companies—really do want testing information to be available to teachers and principals for diagnosis and planning. The common solutions, however, either give raw scores for subscales (to avoid having to report uncertainty and errors) or make little copies of the state test administered throughout the year. These latter tests might be called *benchmark tests* or similar names, but they really promote a micro-summative perspective in classrooms. The development process uses content standards to build summative assessments, which, in turn, dictate the content of formative assessments.

This process creates the practice everyone deplors—teaching to the test. Moreover, much of the curricula sold today reflect reverse engineering by constructing (or, at least constraining) the curriculum on the basis of the test. It seems that curriculum developers are saying, “Let’s figure out what we can do in the curriculum that can help answer the items on the test.”

We need to reverse this, using better forms of content standards to build good formative assessments, which then become summative assessments. Teachers and students in classrooms should be the deciding factor about what is good for instruction. We can then take samples of the formative assessments and create summative assessments in a systematic way that really does reflect the curriculum that is being taught in the schools.

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.

How to Make This Happen

Good instructional assessments are impossible as long as the curriculum in U.S. education is *a mile wide and an inch deep*, which has become the classic description of math and science teaching and applies to other disciplines as well. An alternative would:

- Encourage a long-term view of student growth that connects standards into educationally meaningful learning sequences.
- Use this strategy, along with developing more concrete examples of what these improved standards mean in the classroom, to make standards clear and accessible to teachers and policymakers.
- Use this strategy to develop more efficient ways to use item information than by testing every standard with multiple items. We cannot get good instruction if teachers have to teach 60-100 objectives and standards every year in every subject.

Using Learning Performances and Progressions to Focus the Content

Learning performances, as described in *Systems for State Science Assessments* (Wilson & Bertenthal, 2005), turn content standards from vague statements to something real because they are “...a strategy for elaborating on content standards by specifying what students should be able to do when they have achieved a standard” (p. 3). *Learning performances* should be a required component of standards documents. *Learning progressions*, on the other hand, tell us where the learning is going, while still recognizing that more than one path leads to competence. They make political decisions about content possible by creating meaningful chunks of content that can be publicly debated (e.g., when the major instruction on evolution should take place) instead of allowing some backroom committee to make the decisions.

One approach to building learning performances and learning progressions is to use *progress variables* (Wilson, 2009a). These are detailed assessments based on a cognitive model and operate at the classroom level. Progress variables combine what is known about learning development with knowledge about how items get more difficult, so we can make the interpretation of assessment results more efficient and useful. The psychometrics of selecting from easier and more difficult items (harder items can help measure easier standards as well as the harder ones) reduce the need for many assessment items.

An Example: The BEAR Assessment System

The BEAR Assessment System, published in *Constructing Measures* (see Wilson [2009b] for a summary, or Wilson [2005] for a more complete account), uses learning performances and learning progressions in a system with four principles and accompanying building blocks:

- *Principle 1:* Assessment should be based on a developmental perspective of student learning; the building block is a construct map of a progress variable that visualizes how students develop and how we think about their possible changes in response to items.
- *Principle 2:* There must be a match between what is taught and what is assessed; the building block is the items design, which describes the most important features of the format of the items—the central issue though is how the items design results in responses that are related back to the levels of the construct map.
- *Principle 3:* Teachers must be the managers of the system, with the tools to use it efficiently and effectively; the building block is the *outcome space*, or the set of categories of student responses that make sense to teachers.
- *Principle 4:* There is evidence of quality in terms of reliability and validity studies and evidence of fairness; the building block is a measurement model that provides for multidimensional item responses and links over time, both longitudinally within cohorts and across cohorts.

The Potential

Many examples of the use of such an assessment system show that we can design good assessments that work in the classroom and that we can build good summative assessments from them (e.g., Briggs, Alonzo, Schwab, & Wilson, 2006; Claesgens, Scalise, Wilson, & Stacy, 2009; Kennedy & Draney, 2007; Wilson & Sloane, 2000). Learning progressions can have a positive impact on teaching and how learning occurs in classrooms, because they tell teachers what they are doing and why. At the moment, teachers do not have ways to understand that, because the curriculum is not structured in a way that helps them to be effective teachers. One of the reasons is because assessments are not doing the job we want them to do.

References

- Briggs, D., Alonzo, A., Schwab, C., & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. *Educational Assessment, 11*(1), 33-63.
- Claesgens, J., Scalise, K., Wilson, M., & Stacy, A. (2009). Mapping student understanding in chemistry: The perspectives of chemists. *Science Education, 93*(1), 56-85.
- Kennedy, C. A., & Draney, K. (2007). Interpreting and using multidimensional performance data to improve learning. In X. Liu & W. J. Boone (Eds.), *Applications of Rasch measurement to science education*. Chicago, IL: JAM Press.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.
- Wilson, M. (2009a). Measuring progressions: Assessment structures underlying a learning progression. *Journal for Research in Science Teaching, 46*(6), 716-730.
- Wilson, M. (2009b). Assessment from the ground up. *Phi Delta Kappan, 91*(1), 68-71.

Wilson, M., & Bertenthal, M. (Eds.). (2005). *Systems for state science assessment*. Washington, DC: National Academies Press.

Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education, 13*(2), 181-208.

For More Information

For more information on this subject, please see the paper by Dr. Wilson:

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