New Assessments, New Rigor

Researching. Synthesizing. Reasoning with evidence. The PARCC and Smarter Balanced assessments are setting their sights on complex thinking skills.

Joan Herman and Robert Linn

Extensive research demonstrates the principle, *What you test is what you get*. Study after study shows that teachers tend to focus on tested content and formats and to ignore what’s not tested (Herman, 2004). This is a prime rationale for the United States’ investment in the Partnership for Assessment of Readiness for College and Career (PARCC) and Smarter Balanced Assessment consortia—to develop assessment systems that will embody the Common Core State Standards, focus schools on supporting the deeper learning required for college and career readiness, and help U.S. students become more competitive with those in the highest-performing countries.
So how are the consortia doing? Our analysis (Herman & Linn, 2013) provides some clues.

**The Major Claims**

Both consortia are using a transparent, evidence-centered design approach (Mislevy, Steinberg, & Almond, 1999) that views assessment as a process of reasoning from evidence—student test responses—to substantiate specific claims about student competence. Think of the claims as the major competencies a test is designed to address (and on which score reports will be based) and, likewise, as the major targets for classroom teaching and learning.

Here’s a summary of PARCC and Smarter Balanced claims in English language arts:

- **Reading**: Students can independently read and closely analyze a range of increasingly complex texts.
- **Writing**: Students can produce well-grounded and effective writing for a variety of purposes and audiences.
- **Research**: Students can build and present knowledge through research and the integration, comparison, and synthesis of ideas.

Likewise, here’s a summary of PARCC and Smarter Balanced claims in mathematics:

- **Concepts and Procedures**: Students can explain and apply mathematical concepts and procedures and carry out mathematical procedures with precision and fluency.
- **Problem Solving**: Students can solve a range of complex, well-posed problems in pure and applied mathematics.
- **Communicating/Reasoning**: Students can clearly and precisely construct viable arguments.
- **Modeling and Data Analysis**: Students can analyze complex, real-world scenarios and construct and use mathematical models to interpret and solve problems.

The claims serve as the basis for developing closely aligned items and tasks. Each step of the process builds on the previous one. If assessment intents—the Common Core standards, claims, and evidence targets—are not well reflected in earlier stages, the final operational test will be flawed.

**Measuring Depth of Knowledge**

We’ve been using Norman Webb’s depth-of-knowledge framework (Webb, Alt, Ely, & Vesperman, 2005) to monitor how well the consortia are incorporating the intent of the new standards and their more rigorous learning goals. The framework defines four levels of depth of knowledge that may be embodied in any given assessment task or item:

- **Level 1 test items** draw on basic knowledge and rote learning—for example, literal comprehension questions in reading or simple one-step word problems in math.
- **Level 2 test items** require some application of what’s been learned and some cognitive processing—for example, finding the main idea of a story when that idea is not explicitly stated or doing a two-step word problem.
- **Level 3 test items** require the student to research, synthesize, reason with evidence, and communicate effectively. For example, an assessment item might ask a student to read an editorial on nuclear energy and use evidence from the editorial to analyze the strength of the author’s argument. In mathematics, students might be asked to make and justify an investment decision on the basis of their interpretation of complex data. At this level, we see the increased rigor of the new standards.
- **Level 4 test items** require extended planning, research, and problem solving that call on students’ self-management and metacognitive skills. For example, students might be asked to research a topic from multiple perspectives and present their findings orally and in writing, using multiple media.

In mathematics, students might use their mathematics knowledge to research and recommend the most cost-effective plan for solving an authentic problem, like building a new structure or buying a used car.

**What We’ve Learned**

In our analysis, we found that the guiding claims convey rigorous academic learning goals and reflect Levels 3 and 4 in the depth-of-knowledge framework. The claims are striking in the attention they give to student capabilities that current state tests typically fail to address—particularly the third English language claim, which focuses on research and synthesis, and the third and fourth claims in mathematics, which focus on reasoning.
communication, and nonroutine real-world problem solving. The consortia are clearly after higher-order thinking skills, but unlike in days past, those skills are not divorced from content. Instead, the new standards and the consortia assessments of those standards fully integrate content with higher-order thinking.

Both the PARCC and Smarter Balanced assessments feature technology-enhanced items as well as extended-performance tasks that open up new possibilities for assessment. For example, rather than simply selecting a response, students may be asked to construct graphs, fill in tables, highlight evidence that supports a point of view, use multiple representations, and construct answers to problems that are at Level 3 of the framework.

The performance tasks reach even further, to Level 4. For example, a sample 7th grade PARCC English language arts performance task asks students to read three texts that describe Amelia Earhart’s bravery. Students must then write an essay that analyzes the strength of the arguments presented in at least two of the texts and that uses textual evidence to support their ideas. A sample 6th grade math performance task asks students to recommend which of three field trip options the class should take on the basis of such data as an actual survey of their classmates’ preferences and a comparison of costs in terms of transportation time and expense. Students must then use available data to justify their decision.

**How Do the Assessments Stack Up Against State Tests?**
A recent RAND study (Yuan & Le, 2012) suggests that current state assessments lack such rigor. Drawing on released items and test forms from the 16 states thought to have the most rigorous tests, Yuan and Le found a preponderance of items at the first two levels of the depth-of-knowledge framework. Only about one-third of the few constructed-response items offered were at Level 3; fewer than 10 percent were at Level 4. In math, Levels 1 and 2 predominated, even for the constructed-response items.

The consortia expectations will be a dramatic step forward in rigor. Figure 1 shows the percentage of Smarter Balanced test items at each of the four levels. In contrast to most state tests’ current meager representation at higher levels of the depth-of-knowledge framework, these figures suggest that more than one-third of their new assessments will be composed of items and tasks at Levels 3 and 4. The situation is similar for PARCC, which has specified that one-third of its items and tasks should be at the equivalent of Level 3 or higher (E. Dogan, personal communication, September 27, 2013).

**What Next?**
We end where we started: If what you test is what you get, then the consortia tests are likely to set a high bar for academic rigor and provide a challenging target for classroom teaching and learning. The increased rigor embodies the intent of the Common Core State Standards and the desire for students in the United States to be internationally competitive and prepared for college and career. At the same time, the demands of these tests are likely to come as a shock for teachers and students alike, as the results of early standards-aligned tests in Kentucky and New York have already demonstrated. Being forewarned is being forearmed.

It’s easy to get lost in the details of targets and depth-of-knowledge levels. Moreover, history suggests that a standard-by-standard approach to teaching and learning does not work. Instead, our advice, based on research on learning, is to focus on the big ideas of what students are expected to
accomplish, the major claims about student learning that the new tests seek to substantiate. Consider how specific standards and evidence targets can be integrated in the development and demonstration of these major competencies through appropriate performance tasks.

Finally, it’s worth underscoring that our analysis is based on consortia plans. Operational tests will be fielded in about a year. Time will tell how well current ambitions come to fruition, a situation we’ll continue to monitor.

References
Yuan, K., & Le, V. (2012). Estimating the percentage of students who were tested on cognitively demanding items through the state achievement tests (WR-967-WFHF). Santa Monica, CA: RAND.

Joan Herman (herman@cse.ucla.edu) is Director Emeritus and currently senior research scientist at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at the University of California at Los Angeles. Robert Linn is Distinguished Professor Emeritus of Education at the University of Colorado at Boulder and co-director of CRESST.