



Supplement to August 2010: Vol. 5, No. 4

YOUR ACCESS TO THE REGION

FEATURE ARTICLE:

AN OVERVIEW OF ASSESSMENT CONSORTIA

By Nick Pinchok, Illinois State Manager

Earlier this year, the U.S. Department of Education released the “Race to the Top Assessment Program” application. As part of the larger American Recovery and Reinvestment Act of 2009, this additional Race to the Top (RTTT) grant makes \$350 million available and seeks to fund a minimum of three assessment consortia to help develop the next generation of summative assessments. In the grant, the U.S. Department of Education is asking that consortia develop both “Comprehensive Assessment Systems” (up to two grants worth \$160 million each will be awarded for this category) and “High School Course Assessment Programs” (one award worth \$30 million will be awarded in this category) that “are valid, support and inform instruction, provide accurate information about what students know and can do, and measure student achievement against standards designed to ensure that all students gain the knowledge and skills needed to succeed in college and the workplace” (U.S. Department of Education, 2010).

These requirements pose a very interesting and complex challenge to the testing community and other key stakeholders to quickly improve the types of assessments that are used in an accountability system, aligned to new standards that have just been released (in English/language arts and mathematics), the types and quality of decisions that will be made using these next generation of assessments, and to maintain and ensure psychometric quality and rigor.

Assessment reform has been foreshadowed, however, for some time. Summative assessments for accountability purposes, required under the No Child Left Behind Act (NCLB), have been criticized by multiple stakeholders for their narrow view of teaching and learning. Under the current system, educators often lack timely access to the scores, rendering them useless for instructional purposes. In a fairly unprecedented action, President Obama asked the education community during his campaign to move beyond a system that forced teachers to “spend the academic year preparing students to fill in bubbles on standardized tests” and toward a system with assessments where students are challenged to perform more difficult tasks and “evaluate higher-order skills” (Obama & Biden, 2008).

Originally, six assessment consortia developed in anticipation of these grants to both develop plans for what next-generation assessments should look like and to fulfill the RTTT requirement that states be engaged in and supportive of consortia that would develop assessments aligned to the Common Core State Standards (CCSS) to earn favor during the first and second rounds of RTTT competition. During the past several months, the six consortia have been reduced to three—the SMARTER Balanced Assessment Consortium (SBAC), the Partnership for Assessment of Readiness for College and Careers (PARCC), and the State Consortium on Board Examination Systems (SCOBES). Each applied prior to the June 23 deadline, and links to their applications can be found at the end of this article.

Strengths, Similarities, and Differences

There are some common philosophical and design aspects of these proposed systems that should lead to more coherent, instructionally sensitive real-time assessments. Some of these shared highlights are:

- Develop fully aligned assessments to the Common Core State Standards.
- Produce a range of sophisticated data compiled from a variety of assessment formats.
- Use technology to improve delivery and scoring efficiency.
- Incorporate summative assessment components during the school year.
- Employ universal design principles.
- Incorporate longer performance-based assessment (PBA) tasks as part of the system.
- Provide access to complementary formative assessment tools.

The SBAC and PARCC applications have many similar concepts and goals. Only SCOBES truly broke from the norm to pursue a more unique application. SCOBES proposes that rigorous, proven college- and career-ready assessments already exist overseas. Proven board examinations, such as those in the United Kingdom and in Singapore, have the developmental history, research base, and standards/curriculum/instruction/assessment coherence that we are seeking through these grants. By taking the best of these systems and modifying them slightly to the CCSS, we can save time and tap into the track record of these systems for preparing high school students for both career-ready certifications and colleges and universities of all levels, especially the most prestigious. SCOBES, born out of and capitalizing on the similar work and research done by the National Center on Education and the Economy, hopes to bring this internationally benchmarked system to the United States quickly if awarded this grant and a similar Investing in Innovation (i3) grant they submitted.

Although many similarities exist, the two largest consortia differ in some underlying principles. One aspect is SBAC's intent to engage and involve teachers in the summative assessment development process more thoroughly. This has been done overseas and is a hallmark of Dr. Linda Darling-Hammond's mission in moving toward internationally benchmarked assessment systems that are "part of a tightly integrated system of standards, curriculum, assessment, instruction, and teacher development" (Darling-Hammond, 2010, p. 3). PARCC also will engage new staff in the item development process, but that will be faculty from institutes of higher education to assist with designing new high school tests. The PARCC program intends to use "through course," mostly likely multiple-choice and constructed-response assessments, delivered at specific times during the year to measure 25 percent and 50 percent of instruction to give teachers "actionable information about student learning on a more frequent basis." SBAC plans to administer two computer-adaptive assessments during the year to measure growth and make available other interim/formative options, including performance-based assessment (PBA) tasks "available for state/local use." In addition, SBAC plans to give anywhere from two to six longer "performance events" (approximately one to two class periods in length) in each content area in Grades 3–8 and high school as part of their summative system to provide their more balanced, multiple measures approach. SBAC also plans to explore longer, richer PBA tasks and will look to individual states to pursue these options initially. PARCC intends to give longer performance-based tasks but only one per grade level. Again, SCOBES' focus is exclusively for high school and will use a variety of assessment formats and tasks to measure standards.

The development of computer-adaptive testing (CAT) and PBA as part of a more formative, pre- and post-assessment summative system is a fairly ambitious departure from our current systems. During the year, rather than at one time of the year, the collection of scores from multiple types of assessments will be combined to provide a wider range of assessment data in the eventual summative scoring. The development of and use of PBA involves more real-time, formative assessment delivery and judgment based on the teacher and actually is "embedded" in the instructional delivery, which, again, is a departure from what teachers, administrators, and the public are used to in the current testing and accountability structure. At the Council of Chief State School Officers (CCSSO) National Conference on Student Assessment in Detroit in June, Oregon, Idaho, Utah, and other states presented on their challenges and

successes with using technology and CAT testing (Oregon has been using this system longer than any other state). These states have clearly led the way for moving past many of the proposed concerns and can be leaders in rolling out more timely and formative assessment data on large-scale assessments for all students, teachers, parents, and other stakeholders. These new forms of assessment have significant student growth, teacher judgment, professional development, evaluation, and accountability ramifications that need to be worked out.

Assessment Perspectives

I reached out to a few assessment leaders who wear state education agency (SEA) leadership, psychometric and assessment research hats for their perspectives on what the next generation of high-stakes assessments should look like and accomplish.

When asked, *“What do RTTT assessments need to do differently than our current high-stakes assessments?”* Stan Heffner, Associate Superintendent—Curriculum and Assessment at the Ohio Department of Education (one of the few states participating in both consortia) said, “These assessments need to provide multiple measures of students' progress at different times throughout the school year and in various ways so that the tests yield multidimensional views of student work, inform good teaching and learning decisions, and capture a more accurate picture of how students are meeting learning expectations of the content standards.” Denny Way, Senior Vice President, Psychometric and Research Services at Pearson said, “A primary goal of the RTTT assessments is to provide more balanced assessment information. The summative assessments that have been used for NCLB are good at providing indications of student outcomes, but they are quite limited in providing information that more directly supports teaching and learning.”

When asked, *“What do you think the RTTT assessments should look like?”* Margaret Heritage, Assistant Director for Professional Development at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) said that they should “(e)mbody learning practices—mirror effective instruction—(and) have a variety of knowledge representations... not just multiple choice.” Way said that they should “include performance-based tasks and portfolios... (and) capitalize on the potential of technology both to administer and score new types of tasks and item types.”

When asked, *“What will be the biggest challenge in developing and implementing the next generation of assessments?”* Heffner said, “Teachers will need both time and professional development to see how the new set of assessments impacts classroom decision making and how the results will provide powerful tools to guide instruction. Changes in practice will be necessary in order to take full advantage of the new standards and assessment systems, and teachers will need resources and help in aligning their instruction. These supports are critical for a successful transition to the new systems.” Heritage said, “Time and money... Is there really enough money to develop innovative assessments(?)... I worry that these two factors will mean that in the end things won't change much.”

When asked, *“What testing challenge do we need to break through and solve to make these new assessments better for students and teachers?”* Way said, “A specific challenge is in effectively engaging students and teachers in the administration of RTTT curriculum-embedded assessments.” Heffner said, “We need to set conditions that reinforce the value of assessment results as barometers of student progress rather than as ends unto themselves. In addition, the assessments employ technologies that most students use every day but may not encounter during school hours. A transition in use of technologies in a student's daily learning (and one that aligns with the modalities of the new assessment system) is a critical challenge that must be overcome—and will be.”

Challenges

Cost

The federal government is making only \$350 million available for the four-year development and rollout of these new, ambitious assessments. This funding literally is not enough to support full content coverage for item development as well as the infrastructure, professional development, scoring, and other

characteristics of current and proposed systems. It is a start, however, and funding for innovative assessment development is an opportunity that the assessment community appears to be pursuing thoughtfully and boldly, based on the applications. Acting as consortia, too, will provide cost savings, and the SCOPE group commissioned by Stanford University recently identified some significant cost savings that SEAs can tap into by using more state-of-the-art processes and acting collaboratively (Topol, Olson, & Roeber, 2010).

New Psychometrics/Comparability

Concerns have already arisen about many new testing and psychometric issues that will emerge soon, such as teacher-scored assessments, comparability, and computer-adaptive testing. After complaints about the limitations of mostly multiple-choice, one-time-per-year testing during the past several years, there is hope that the proposed, more student- and teacher-friendly assessments can more positively impact instruction and learning but could be derailed by some current psychometric challenges, especially concerning teacher-scored PBAs. Some questions that will need to be addressed are:

- What can and should we accept as acceptable levels of reliability, validity and comparability?
- What are the breakthrough technologies or newer statistical analyses that we should adopt or pilot?

Three testing companies—ETS, Pearson, and the College Board—recently published an article (Lazer, et al., 2010) addressing some of these issues. Fortunately, they proposed multiple scenarios where varying levels of comparability could be maintained and analyzed across consortia, from using common items at one end of the spectrum to “scale anchoring” for empirical analysis of proficiency levels with no similar items at the other end. Most assessment systems overseas are not explicitly developed around our priority of one-time, high-stakes accountability, and we clearly can learn from the high-quality judgments they make on more instructionally sensitive, teacher-scored assessments as part of our retooled accountability and psychometric models. While doing research on large-scale, PBA systems overseas, Dr. Terry Crooks of New Zealand’s National Education Monitoring Project said, “Whatever we try to do in assessment should always begin with the question: What is this going to do to motivate, inspire, and guide teachers and learners towards best practice, and what are the risks that it will do the opposite? How do we try to ensure the benefits while minimizing the damage?” There is a way through this, and our best testing development and psychometric minds will be hard at work during the next few years figuring it out.

Professional Development

As these more balanced assessment systems develop, become implemented, and garner higher stakes, the requirement for comprehensive professional development will clearly become a priority. Moving from assessments that are passed out and collected to be scored externally to more embedded as part of instruction and scored by teachers is a fundamental shift. Coordinated efforts to provide these supports between a consortium, their participating SEAs, their regional educational service agencies, districts, and other partners in this process will need to emerge. Some international systems with more balanced, large-scale systems have implemented quality assurance systems that could be models for the United States, such as Queensland (Australia) and Hong Kong. Sophisticated analyses for more performance-based and student growth-oriented assessments, and the pending teacher and administrator evaluation systems based on them, however, will be unique challenges we experience.

Conclusion—Getting It Right

All of this sounds good and should lead to better high-stakes assessment that lead to better instruction. In a recent joint paper by the National Governors Association (NGA) and CCSSO (2010), they state that there is “an unprecedented opportunity for states to work together to dramatically to improve the quality, cost-effectiveness, and comparability of state assessments.” As the Race to the Top Assessment Program competition seeks the “next generation” of higher quality assessment programs and approaches (i.e., performance assessments, computer-adaptive assessments), there is the chance that quality, cost-effectiveness, and comparability will not improve incrementally at the “right” levels. That may not be a bad thing, though, if these assessments lead to higher levels of cognition at the student level and more engaged, project-based instruction at the teacher level. It is an exciting and challenging time in the

assessment world, though, and we'll wait to see what happens after the \$350 million awards are handed out.

Links to Assessment Consortia Applications

SBAC <http://www.k12.wa.us/SMARTER/>
PARCC <http://www.fldoe.org/parcc/pdf/apprtcasc.pdf>
SCOBES <http://www.edweek.org/media/rtttnarrative%20final.pdf>

References

- Darling-Hammond, L. (2010). *Performance counts: Assessment systems that support high-quality learning*. Retrieved August 2, 2010, from http://flareassessment.org/resources/Paper_Assessment_DarlingHammond.pdf
- Lazer, S., Mazzeo, J., Way, W. D., Twing, Jon., Camara, W., & Sweeney, K. (2010). *Thoughts on linking and comparing assessments of common core standards*. Retrieved August 2, 2010, from http://www.ets.org/Media/Home/pdf/14518_RTTTA_WP_WEB.pdf
- National Governors Association & Council of Chief State School Officers. (2010). *Designing common state assessment systems*. Washington, DC: Author. Retrieved August 2, 2010, from <http://www.nga.org/Files/pdf/1004NGACSSOASSESSMENTS.PDF>
- Obama, B., & Biden, J. (2008). *Barack Obama and Joe Biden's plan for lifetime success through education*. Retrieved August 2, 2010, from <http://www.barackobama.com/pdf/issues/PreK-12EducationFactSheet.pdf>
- Topol, B., Olson, J., & Roeber, E. (2010). *The cost of new higher quality assessments: A comprehensive analysis of the potential costs for future state assessments*. Stanford, CA: Stanford Center for Opportunity Policy in Education. Retrieved August 2, 2010, from http://edpolicy.stanford.edu/pages/pubs/pub_docs/assessment/scope_pa_topol.pdf
- U.S. Department of Education. (2010). *Race to the Top Assessment Program*. Retrieved July 30, 2010, from <http://www2.ed.gov/programs/racetothetop-assessment/index.html>