

Getting Student Assessment Right

by

Gregory J. Cizek, PhD

Guy B. Williams Professor of Educational Measurement and Evaluation
University of North Carolina-Chapel Hill

February 8, 2016

As all states wrestle with how to implement the flexibility afforded under the *Every Student Succeeds Act* (ESSA), it is critical to focus on the following five fundamental aspects to ensure a high quality assessment system.

1) Ensure rigorous content standards are in place.

The Common Core State Standards (CCSS) are one example of content standards that embody high expectations for what students should learn in English language arts and mathematics—but they aren't the only example. Many states that have not adopted the CCSS, or that have partially adopted the CCSS, have developed their own challenging content standards for all students. Within the next two years, it is likely that the CCSS will need to be revised; all sets of content standards—such as the North Carolina Standard Course of Study—are routinely reviewed and changes made.

First, it is essential that the content standards are adopted in North Carolina be vetted by three major constituencies: educators, college and career stakeholders, and parents. The North Carolina learning expectations must be grade-level appropriate for students, challenging, and provide all students with the knowledge and skills they need to be successful at the college or career option to which they aspire.

Second, it is unlikely that a generic national assessment program and strictly multiple-choice testing will provide adequate alignment to whatever challenging content standards are ultimately adopted. Although commercial test suppliers might claim their products are aligned to North Carolina's content standards, the reality is that the commercial vendors create one-size-fits-all products that are as reasonably aligned to as many states' sets of content standard as possible. North Carolina must demand an assessment program worthy of the relevant and rigorous content standards deemed appropriate for our students. The state must also demand testing formats that measure those challenging content expectations. For example, only constructed-response formats (e.g., essays, oral presentations, short answer questions) are appropriate if it is decided that students must demonstrate critical thinking skills, such as the ability to analyze written information such as newspaper articles and develop their own reasoned positions on issues. Multiple-choice items surely have a place in a statewide assessment program for their ability to cover more content standards and do so more efficiently and reliably, but it is not possible to measure next-generation skills with multiple-choice items alone.

2) Keep a focus on purpose...

Tests are like tools: each one is designed for a specific purpose. It is surely *possible* to drive a screw into a wooden board with a hammer, but the results will often be less than satisfactory. Likewise, assessments have different purposes that they are each designed to accomplish.

One purpose that all states have is to give students, parents, and educators a good idea of each student's overall achievement in subjects like math, reading, and science. At the same time, state legislators and policymakers are typically interested in monitoring the progress of the entire educational system, in ensuring the responsible use of education dollars, and in evaluating the effectiveness of various statewide educational interventions and programs. Relatively short, end-of-year assessments such as End-of-Grade and End-of-Course tests accomplish these

purposes fairly well. They are not, however, designed to provide fine-grained information to students or teachers; they are not the appropriate tool to diagnose individual student's strengths and weaknesses; they are limited in the information they can provide for evaluating educators' effectiveness. Different tools are needed for those purposes.

For example, specially designed *formative assessments* can be developed and implemented to gauge where students need the most assistance and to inform teachers about the appropriate next steps to take in classroom instruction. There is no reason that these assessments must be "standardized," contain specified item formats, or have results reported to any level beyond the classroom or school. The key point here is that a primary consideration in any assessment system must be identification of purpose; every component of a coherent assessment system must be developed with steadfast fidelity to the unique purpose it is intended to serve.

3) ... and an eye on technology.

Technological innovations have great promise for assessment systems. For example, computer delivery of large-scale tests is not only currently possible, but allows for more authentic types of technology-enhanced items and tasks to be presented to students, for efficient keyboard entry of automated scoring of essay responses, and advantages in terms of cost savings (e.g., printing, shipping) and test security.

But technology also has limitations. For one, computer-based testing (CBT) rarely translates into reduced testing time. It typically takes the same amount of time to get dependable information about student learning from a paper and pencil test as it does from a CBT. If next-generation assessments are developed to require students to engage in more critical thinking and analysis (*see Issue #1*), it is likely that they might even take more time. Current technology in terms of score reporting is also limited. For example, some proposals for "through course" assessment programs have included the possibility of interim assessments at designated time points through the school year, with a final summative assessment at the end. However, there is presently no accepted technology for how to combine these results into a single, valid index.

To illustrate the challenge, let us imagine tests as described above given at four key points during a school year; for convenience, let's label the four tests as T1, T2, T3, and T4. Let us also imagine a third grade student who struggled mightily with learning how to add fractions such that the student scored very poorly on T1 (the test that covered that topic in the mathematics curriculum) but the student eventually demonstrated mastery of adding fractions on T4, which was a cumulative test that sampled topics from across the full school year. How should this student's performance be reported? It would seem wrong to simply average the students' scores when the T4 result is the only one that accurately represents the student's mastery of the fractions. Or, if we T4 was *not* a cumulative test, but only covered the last portion of the school year, the student would not get "credit" for his or her eventual mastery. The bottom line is that assessment in schools must be kept in line with current technologies—regardless of whether the technology relates to computers or score reporting—so that the results of testing make sense.

4) Accountability works, if it's done right.

As an educator in a university setting, I can attest that the accountability mechanism of "publish or perish" functions precisely as intended, motivating faculty focus on their scholarly productivity. Similarly, research has demonstrated that accountability policies for K-12 contexts typically have the desired effect of focusing effort on the student achievement. Two factors must be balanced, however.

First, to function as intended, accountability systems must include diverse elements and have real consequences. The days of accountability systems that monitor only student test score gains are—or should be—long gone. Other factors that might be included in a next-generation accountability systems are the extent to which a school stimulates parental engagement, the extent to which the school is successful in getting students to enroll in challenging learning experiences such as Advanced Placement (AP) courses, the school's success in improving the safety of the learning environment, increasing the rates at which students graduate on time or engage in community service, or myriad other valuable educational outcomes.

Second, whereas we know that accountability systems without consequences aren't effective, we also know that one-sided accountability is both ineffective and unfair. The term *symmetry of consequences* has been introduced to refer to accountability systems, where all involved have some "skin in the game." For example, accountability systems where only educators or school systems face sanctions for poor student achievement are often—and rightly—judged to be unfair to educators. Systems where only students face penalties for poor performance also seem unfair. The key policy goal should be to create accountability systems that balance fair, clear, and reasonable consequences for all involved.

5) Demand assessment literacy.

Finally, it is an embarrassment that formal training for educators in assessment is so spotty. For example, such training is not required to become a licensed teacher, principal, even superintendent, in the state of North Carolina. It is hard to imagine that fundamental competence in assessment is not mandatory in today's environment where assessment information is so prevalent, so necessary for ensuring student success, and so fundamental to sound, evidence-based decision- and policy-making at all levels. Failing to equip pre-service teachers and administrators with the essential grounding in assessment guarantees that educators will not have the skills to develop the kinds of tools they need (e.g., formative assessments); they won't be equipped to interpret assessment results and translate those results into instructional practice; and they will continue to view assessment and policy initiatives that include unfamiliar assessment components with skepticism. Assessment literacy must be a component of whatever assessment systems is contemplated for the future.

Overall, the coming next generation of content standards, assessments, and reporting mechanisms all offer great opportunities to improve learning for all students. Of course, there are ever-present opportunities to miss these chances to help students acquire and demonstrate the knowledge and skills they will need for success in their chosen postsecondary paths. Thoughtful, coherent, planned assessment *systems* are surely possible, but only if they are rigorous, maintain focus on the unique purpose of each component of the system, balance technological opportunities and limitations, and include initiatives to develop effective accountability and enhance assessment literacy.