

Realistic Expectations of Accountability Assessments:

One Test Can't "Do It All"

Stuart Kahl | March 2016

I wrote a piece, titled "Enough Is Enough!" that appeared in the April 2, 2009 edition of Education Week. It began with a fictitious quote that mimicked real campaign messages from gubernatorial candidates in one state:

"Let's replace our current accountability assessment with a single, summative, formative, adaptive, diagnostic, general achievement test that measures growth and yields immediate results teachers can use right away to modify their instruction."

Despite the absurdity of this thinking, many states and local education agencies, while seeking testing contractors, were issuing (and continue to issue) requests for proposals (RFPs) reflecting similar expectations of their tests.

The mindset suggesting that tests can serve numerous purposes and solve any and all major education-related problems has no doubt contributed to today's overuse and misuse of tests.

Compounding the problem, the quick turnaround of results on efficient, computer-delivered tests has further nurtured teachers' expectations of statewide accountability tests, which teachers claim are useless to them unless they can see the results almost immediately. Unfortunately, the desire for inexpensive machine-scoring for quick turnaround of results means that these tests often emphasize low-level knowledge and skills, with a resulting negative impact on instruction and student achievement.

Educators and policy makers alike seem to think that a 50-item, predominately machine-scoreable, general achievement measure addressing a thin sampling of the material covered in a full-year course can provide diagnostic information on individual students' needs and tell teachers what they need to do the next day. If teachers in a school must count on a state test for information about an individual student's capabilities with respect to specific curricular content being taught around the time of the state testing, then the school has larger problems than just performance on the state test. A total test score and a few subtest scores are the most those tests typically can support.

“ . . . quick turnaround of results means that these tests often emphasize low-level knowledge and skills . . . ”

Student and School Accountability

Years ago, the emphasis of state testing was—appropriately—on school results. As the stakes associated with those results ratcheted up for educators, there were growing concerns that educators were being held accountable for student test performance, while the students themselves were not. The students were not motivated to do their best on the state tests. Consequently, many programs began producing reportable individual student results. Holding students more accountable for their test performance and NCLB's holding schools accountable for **all** their students are good things.

Best Uses of State Accountability Test Results

As in the past, a real value of state accountability testing is the utility of its results for instructional program evaluation and improvement. A school's test results should raise questions, rather than only answer them. Why did we perform poorly in this area of mathematics? Why did this group of students score lower than that group of students? These questions require investigation, and their answers can inform programmatic improvements. For this purpose, quick turnaround of results is not necessary. In fact, many programmatic changes should probably be made based on patterns of results that show up over multiple years. Thus, extended performance tasks, which can tap higher-order skills, but require time for human scoring, can play a more significant role in accountability testing.

“ . . . two parts of the ESSA give states the flexibility to implement innovative programs . . . ”

Today's Opportunity: Balanced Assessment Systems

As part of school reform efforts, student testing, which remains a critical and integral component of instructional programs, is undergoing significant transformation. Let's hope that reform initiatives result in balanced assessment systems that benefit from the good features of past testing and the good features of recent innovative approaches, too. A balanced assessment system strikes appropriate balances between formative and summative assessment, between local and state assessments, and between emphases on foundational skills and deeper learning. All these can be achieved, but not with a single test.

Impact of ESSA and NCLB

Even though the Every Student Succeeds Act (ESSA) includes, in my opinion, some significant improvements in assessment provisions over No Child Left Behind (NCLB), it still perpetuates the idea that a single test can do more than it really can. ESSA requires that the assessments “produce individual student interpretive, descriptive, and diagnostic reports...” that allow different audiences to address specific needs of students and that are provided “as soon as practicable after the assessment is given.” Holy cow! Did U.S. legislators take my *Education Week* “quote” seriously?

Fortunately, state assessment systems that don't accomplish the impossible, and none of them can or do, still pass muster with the U.S. Department of Education. Furthermore, two parts of the ESSA give states the flexibility to implement innovative programs, such as ones that might involve both curriculum-embedded performance assessments and an end-of-year summative measure. First, the ESSA provides greater clarity in the meaning of “multiple measures” through the mention of performance and portfolio assessments. Second, the law allows results of interim assessments to contribute to accountability assessment results. Actually, some innovative components of assessment programs could be designed to satisfy teachers' desire for immediate, useful results.