

COMBATING THE FRAUDULENCE OF STANDARDS-BASED ASSESSMENT*

W. James Popham
University of California, Los Angeles

“Standards-Based Assessment” really sounds quite wonderful. Yet, in most educational settings, it is a flat-out fraud.

In my *Webster Dictionary*, a fraud is defined as “something that is not what it pretends.” Well, any sort of beneath-the-surface look at today’s standards-based assessment will soon reveal that this alluringly labeled breed of testing is simply loaded with artifice.

Re-defining “Standard”

I suppose it all started a decade or so ago when American educators began bastardizing the term *standard*. A “content standard,” it was argued, should describe the skills and knowledge that educators want their students to learn. Yet, although *Webster’s* offers fully 27 definitions of “standard,” not one of those definitions remotely approximates a description of the skills and knowledge that students should learn.

Content standards, as used (misused?) these days, simply refers to what used to be called “curriculum goals” or “instructional objectives.” I’m really not sure precisely when it was that educators began to describe their curricular aspirations as content standards, but I have hunch regarding *why* they did so. “Standards,” in the traditional way that this noun has been used, and especially if preceded by the adjective “high,” becomes an educational entity that simply reeks of goodness. Who, in their right mind, could ever be opposed to “high standards?” So, if the nation’s educators claimed that they were setting out to promote students’ mastery of high standards, who would dare criticize such a laudable aspiration? “High standards” is a phrase that, by definition, alone, elicits applause.

Thus, when the phrase “standards-based assessment” subsequently wormed its way into our educational lexicon, it also possessed ample goodness-reeking qualities. Standards-based assessment was touted as an approach to measuring whether students had mastered all sorts of nifty skills and knowledge, that is, nifty content standards. Moreover, because such an assessment strategy was supposedly based on those nifty content standards, there was a clear implication that standards-based tests would surely assist educators in their efforts to promote students’ mastery of the content standards under consideration.

* A version of this essay, appearing under the title, “Trouble With Testing,” appeared in the February 2003 issue of the *American School Board Journal*, Vol. 190, No. 2, pp. 14-17.

In many instances, for example, a state's curriculum officials first identified a set of content standards for their state, then statewide standards-based tests (either built or bought) were used to ostensibly assess students' mastery of those state-sanctioned content standards.

And that's where fraud frolics onto the scene because (1) standards-based tests typically don't measure the skills and knowledge they purport to measure and (2) most standards-based tests don't, as is claimed, help educators do a better instructional job. Standards-based assessment, clearly, "is not what it pretends."

Pretense Number 1: Measuring Myriad Content Standards

Suppose a state's education leaders have formally approved a set of skills and knowledge that, therefore, constitute the state's official content standards. Suppose further that a state's education officials have installed statewide standards-based tests that are supposed to measure students' mastery of those state-sanctioned content standards. If both these events take place (and they have in about half of our states), then it would be reasonable to assume that the state's curricular aims were, indeed, being measured. That assumption is almost certainly unwarranted.

Much of the problem stems from the enormous number of content standards typically staked out by a state's curriculum specialists. Remember, these curriculum specialists are, in every sense of the term, *specialists*. And most specialists simply adore their fields of specialization. Thus, for instance, when a state-convened panel of 25 mathematics teachers and mathematics curriculum experts is directed to determine what mathematics content the state's students should master, you can safely predict that those mathematics specialists will want the state's students to learn *everything*, that is, everything even remotely mathematical.

And that's why many states have now approved literally *hundreds* of content standards to be mastered by students at given grade-levels. Sometimes a state's curricular architects may have appeared to adopt a much smaller number of content standards, for instance, only a dozen or so fairly broad standards per grade level. However, closer scrutiny of those content standards will typically reveal that beneath each of the dozen broad curricular aims, there are numerous "benchmarks" or "indicators" set forth. And, it is only these more specific descriptions of skills and knowledge that are stated at a level of clarity suitable for the state's teachers to tackle instructionally. As a consequence, there are still way too many curricular aims to teach in a given school year. And, similarly, there are way too many curricular aims to assess in the hour or so typically available for the administration of any standards-based test.

What, then, do those who must construct standards-based tests typically do in this all-too-common situation? The answer is that they *sample*, that is, they design their tests to measure some, but not all, of their state's sprawling collection of content standards. And such a sampling-based solution means, of course, that a good many state-sanctioned content standards are left unmeasured. Yet, the image is frequently

fostered that a state's standards-based assessments actually measure the state's complete array of content standards. That's just not true.

To make matters worse, most worthwhile content standards can't be meaningfully assessed with only one or two items. Yet, that's what usually happens for those content standards that, having come up as winners in the sampling derby, are actually measured. Such less-than-complete measurement is better described as "tokenism" than as "standards-based."

Some commercial test publishers, of course, have claimed that, to a very considerable extent, their "off-the-shelf" standardized achievement tests are already sufficiently aligned with a given state's content standards so that, in fact, a ready-made standards-based test eagerly awaits adoption. In today's high-stakes testing arena, where enormous numbers of dollars find their way to those firms whose tests are adopted, decision-makers must be particularly wary of such alignment claims.

In the field of education, *alignment* is the only 9-letter word that functions like a 4-letter word. This is because the stringency levels of alignment are often "adjusted" when determining whether content standards are suitably addressed by a standards-based test. It is often in the self-interest of test publishers to have a state's potential test-adopters regard a test as being more aligned with a state's content standards than is actually the case. Any alignment study conducted by those who have a special interest in the study's outcome should be regarded as what it is, a special-interest alignment study.

To reiterate, in almost all instances, standards-based tests fail to measure the full set of content standards on which they are supposedly based. And even those content standards that are actually addressed in a standards-based test are usually measured in only a superficial fashion.

Even if certain content standards are assessed in one year's state test, and other content standards are assessed in another year's test, for any particular year there is no satisfactory picture presented of how students are doing with respect to a state's full collection of approved content standards.

It is time for educational assessors to abandon any implication that a state's standards-based tests actually assess a state's too-numerous content standards. To pretend that a few tests, administered in an hour or two, can satisfactorily measure a state's myriad curricular aspirations is little more than assessment hypocrisy.

Pretense Number 2: Supporting Instruction

If a state's content standards set forth the skills and knowledge to be learned by the state's students, then an instructionally supportive standards-based test ought to provide information regarding *which* standards have been well-taught and *which* standards haven't. How else can the state's teachers identify the parts of their

instruction that should either be fixed or are just fine as is? But that's where another pretension of standards-based assessment becomes apparent, because today's standards-based tests do not yield standard-by-standard results, either for districts, schools, or individual students.

Instead, what's usually reported is an overall score for, say, mathematics or reading. Occasionally, there are attempts to provide some sort of subcategory reports, for instance, a student's reading subscores dealing with the student's mastery of (1) fictional or (2) non-fictional reading materials. But standards-based tests currently do not provide teachers, students, or students' parents with the sort of standard-by-standard results with which appropriate instructional decisions can be made.

What *instructional* sense can a teacher make out of a total "mathematics" score, or even a "geometry" sub-score? If standards-based assessment is going to help teachers do a better job of teaching their students, then it must supply teachers with the level of information regarding their students' status that allows teachers to tell which aspects of their instruction are tawdry and which aspects are terrific. Currently, the nation's standards-based tests don't do so.

Today's use of standards-based assessment is akin to what might happen at a horse race, such as the Kentucky Derby, if the race's results were only reported as the *average* finishing time of *all* the horses that ran the race. That is, the race-report would provide no indication of an individual horse's race-finishing time. Sound pretty stupid? Well, what we're doing now with standards-based assessment is every bit as stupid!

But where our current standards-based assessment situation gets even worse (than my absurd horse-racing example), is that many standards-based assessors are loudly claiming that their tests are meaningfully supportive of the educational process. It just isn't so.

Antidotes to Fraudulence

In order to make standards-based assessment a legitimate contributor to improved instruction, we simply need to be honest about what is and what isn't measurable in the modest number of minutes available for the assessment of students. Moreover, so that standards-based assessment will be instructionally supportive, we must decisions *chiefly* on what will be good for instruction. I will close out this brief analysis by offering several suggestions based on the positions taken by the Commission on Instructionally Supportive Assessment in October 2001.*

* Commission on Instructionally Supportive Assessment. (1) *Building Tests That Support Instruction and Accountability: A Guide for Policymakers*; (2) *Illustrative Language for an RFP to Build Tests That Support Instruction and Accountability*. Washington, DC: Author, 2001. Available online at www.aasa.org, www.naesp.org, www.principals.org, www.nea.org, www.nmsa.org.

Prioritized content standards. Honesty demands that educational assessors indicate they can not measure existing sets of content standards in the time available for testing. Thus, assessors must insist that curriculum specialists prioritize their content standards, from the most important to the least important. At that point, assessment personnel must devise tests that can, with reasonable—albeit not perfect—accuracy, assess as many of the top-ranked content standards as possible.

What this means, of course, is that there will be far fewer content standards actually assessed in statewide accountability tests. It might turn out that a statewide test *actually* assesses only a half-dozen or so content standards rather than *allegedly* assessing a half-hundred or so. That being so, in most settings it will be imperative for curriculum and assessment personnel, *working together*, to reconceptualize content standards so that their focus is on broader, yet measurable, skills or bodies of knowledge.

This reconceptualization will be tough work, for there is considerable art required in conceptualizing content standards so that they are both sufficiently clear, sufficiently broad, yet still amenable to cost-affordable assessment. One of the best illustrations of how this can be done is currently represented in the nation's widespread use of writing samples to measure students' composition skills. These writing-sample tests are accompanied by scoring guidelines featuring instructionally addressable evaluative criteria. With one or two student writing samples, it is possible to come up with a reasonably valid inference about a student's compositional prowess. We need to carve out our desired skills and/or bodies of knowledge at similar levels of generality and clarity.

Palatable descriptions of the content standards to be assessed. Once the highest-priority content standards have been identified for assessment, then a succinct, teacher-friendly description of each content standard must be developed and made available to all concerned. These descriptions, written so that they are especially palatable to teachers, must set forth the nature of the skills and/or bodies of knowledge so a teacher can design instructional plans aimed toward the content standard itself, not toward the specific test-items intended to assess that standard.

Standards-based tests, if they are to be instructionally supportive, must provide clear descriptions of the cognitive demands a student will encounter when demonstrating mastery of any to-be-measured content standard. Such descriptions, of course, will permit teachers to stop teaching toward particular items on a standards-based test. There's far less need to teach toward specific items on a test if teachers understand the essence of the skills or bodies of knowledge represented by that test's items.

These brief, readily understandable descriptions will surely require collaborative work from both curriculum and assessment personnel. These descriptions, however, represent a *sine qua non* if a standards-based test is going to be truly instructionally supportive. Shorter versions of these assessment descriptions, provided to both

students and their parents, can also have a decisively positive impact on students' learning.

Standard-by-standard reporting. Another integral feature of an instructionally supportive standards-based test is that its results must be reported on a per-standard basis at the individual student level. Because more items will typically be required for any sort of meaningful standard-by-standard reporting at the student level, this is a key reason why instructionally supportive standards-based tests dare not try to measure too many content standards. From an *instructional* perspective, it is far better to measure five important content standards *well* than it is to measure 25 content standards *badly*.

Standard-by-standard reporting, of course, will make it possible for teachers, students, and parents to take suitable action once it has been determined that particular content standards have not been mastered by a child. Standards-based tests without standard-by-standard reporting will not help students learn what they should be learning. Standards-based assessment without standard-by-standard reporting is crazy!

Classroom assessment of state-unassessed content standards. We want students to learn what they really need to learn, but large-scale standards-based assessments can only measure so much. Accordingly, state officials will typically need to rely on local assessment of other state-unassessed, yet worthwhile outcomes.

Ideally, I would like to see a state, or perhaps a consortium of states, develop assessment instruments suitable for this purpose, then turn over those classroom assessment tools to a state's school districts. In some instances, a state's education officials might stipulate that certain specified content standards, or perhaps a minimum number of state unassessed content standards, must be assessed locally each year. Students' performances on these local assessments could then be reported in the new federally required school, district, and state level school-focused report cards. In other instances, more flexibility could be built into a state's stipulations regarding how state-developed classroom assessments should be used locally.

The essence of this suggestion is to make it clear that, because honest standards-based assessment at the state level will limit the number of state-assessed content standards, a good deal of serious classroom assessment has to be done locally. States, however, need to provide the assessment tools to carry out this kind of classroom assessment. It is unrealistic to expect that already overburdened local educators will be able to generate their own high-quality classroom assessments.

Four fix-it recommendations. The four suggestions proffered above could supply meaningful resuscitation to today's floundering standards-based assessment movement. Those suggestions were: (1) prioritize content standards and measure only those that can be assessed with, for instructional purposes, reasonable accuracy; (2) develop and disseminate palatable descriptions of the skills and/or knowledge being assessed; (3) provide standard-by-standard reports of individual students' status; and

(4) address the measurement of state-unassessed content standards via state-developed but locally administered classroom assessment instruments.

Hunky-Dory Standards-Based Assessment

What would an instructionally supportive standards-based assessment program look like? Well, here's my view—an image that, I believe, is in accord with the assessment/ accountability provisions of *The No Child Left Behind Act of 2001*. In my briefly described vision of a standards-based assessment Valhalla, I have attempted to incorporate optimal instructionally relevant features that can be implemented within the realities of both cost and time.

A state's curriculum and assessment specialists, working collaboratively, would have identified roughly 5-10 highest-priority content standards in both reading and mathematics for each of grades 3-8 and 11. One, two, or three of these content standards in both reading and mathematics would be assessed via some form of constructed-response testing. For instance, at specified grade levels, children might be obliged to read a fictional and/or nonfictional passage, then generate (1) a "from scratch" description of the passage's theme or main idea and (2) answers to questions regarding key details presented in the passage.

Other than a small number of content standards (per subject) designated for constructed-response measurement, the remaining content standards to be assessed would be measured via selected-response items. Typically, because the content standards to be assessed in this manner will be relatively broad in scope, somewhere between 5 and 10 selected-response items will be needed for the assessment of each content standard (so as to permit per-standard reporting).

Statewide tests would be administered in the spring, and results would be supplied to schools by August so that teachers could determine, prior to the new school year, if any of their intended instructional activities for the upcoming academic year would need to be altered. If teachers, during the school year itself, need en route evidence of students' growth toward particular standards' mastery, this information would need to be collected via the timely administration of classroom assessments.

The state would supply districts with (1) classroom assessments suitable for measuring a reasonable number of the state's important but *state-unassessed* content standards and also (2) classroom assessments measuring each of the *state-assessed* content standards. The classroom tests for *state-unassessed* outcomes might be totally optional or, perhaps, a modest number could be required at the local level so students' results could be reported annually. The classroom tests for *state-assessed* outcomes would be totally optional, but could be used by teachers to determine their students' per-standard progress early enough in the school year so there is still time to address students' deficiencies instructionally before the close of school.

It is, to me, patently silly to try to administer statewide tests early enough in a school year, say, in January, so as to provide teachers with evidence of students' mastery prior to a school year's conclusion. Classroom assessments, devised by the state but administered as needed by teachers during the year, seem to be a better way to go. Such classroom assessments can provide teachers with as-needed evidence of their students' progress toward the state's most significant content standards.

Standards-Based Assessment: A Good Idea Gone Awry

If the architects of standards-based assessments remember that their approach to measurement must fundamentally be predicated on its contribution to improved teaching, then there will need to be substantial alterations in today's version of standards-based testing. As matters stand, standards-based assessment is, unarguably, a fraud. But it doesn't need to be.