

ASSESSING MASTERY OF WISH-LIST CONTENT STANDARDS

W. James Popham
University of California, Los Angeles

After characterizing most of today's content standards as wish-list curricular aspirations, the writer argues that to assess such standards properly, educators must prioritize them uncompromisingly. Only those content standards determined to be of the highest priority and also measurable on a per-standard basis should be tested via large-scale assessments. The remaining standards should be measured by classroom assessments or should serve as targets for teachers' instruction.

The first line of a song I often heard during my youth was, "Wishing will make it so." Well, I suspect that the educators who created today's content standards must have been humming and/or whistling that tune when they did so. Because, with few exceptions, today's content standards are most assuredly glorified *wish lists*. The architects of these sometimes interminable lists of content standards were identifying all the good things they'd like students to learn.

But *wish-list content standards*, particularly when we try to determine students' attainment of those standards, create some very serious assessment obstacles. Thus, what I want to consider in this brief analysis is one way that educators can, from an *instructional* perspective, defensibly assess students' mastery of wish-list content standards.

Different Names, Same Games

Content standards, of course, describe the skills and knowledge we want our students to attain. Eons ago, when I was a first-year teacher, we called such things *objectives* or, sometimes, *goals*. But because I always strive to be fashionable, I shall employ the lingo of the current content-standards crowd. Who knows, if I hang around long enough, *objectives* and *goals* may return!

Educators want children to learn the things that need to be learned. After all, that's why most folks aspired to be teachers in the first place. They wanted to help young people master the knowledge and skills that are required during school and, even later, when school is over.

So, decades ago, it was not uncommon for a state's educators to hammer out the particular collection of knowledge and skills it was hoped that state's children would acquire. The resulting documents would often be designated as a state's *curricular framework* or its *curriculum syllabus*. It was hoped that a state's teachers consulted such documents prior to devising their instructional plans.

* This essay appeared in the December 2000 issue of the *NASSP Bulletin*, Vol. 84, No. 620, pp. 30-36.

In my experience, such curricular documents, even though sanctioned by high-level state officials, had almost no influence on the day-to-day instructional activities of most teachers. Curricular frameworks, typically thick documents (because wish-listing was common even back then), rarely left a teacher's bookshelf once placed there.

And even when educational objectives became popular as a way of representing a teacher's instructional aims, those objectives often had scant impact on what went on inside classrooms. During the 1960s, *behavioral objectives*, that is, instructional objectives describing the post-instruction behaviors sought from students, became especially popular with most educational policymakers.

And I was right in there pushing for behavioral objectives as hard as I could push. I wrote books about behavioral objectives, authored articles about behavioral objectives, created filmstrips about behavioral objectives, built a bank of behavioral objectives, and gave gobs of speeches about the instructional dividends available to teachers who aimed their instruction at behavioral objectives. Indeed, the esteemed evaluator Michael Scriven once quipped that I was "the licensed midwife to the birth of behavioral objectives."

But even my love affair with behavioral objectives cooled off after several years. It appears that a temporary fascination with behavioral objectives, just as adolescent acne, is a phase that some educators must go through. I was one of those educators.

I had discovered that many teachers, if requested, could trot out a set of explicit behavioral objectives. But, on further analysis, those objectives had almost no influence on what such teachers were doing in their classrooms. The two big-bopper problems with yesteryear's highly specific behavioral objectives were: (1) there were too many and (2) they were not properly assessed. Because behavioral objectives were far too numerous, they simply overwhelmed teachers. What classroom teacher has time to deal with, *literally*, hundreds of behavioral objectives? Sensible classroom teachers, of course, simply placed their state-approved or district-approved behavioral objectives on the bookshelf (right there alongside some former teacher's dusty curriculum syllabus), then got on with the real-world business of teaching kids.

Another shortcoming of the behavioral objectives movement was that, for the most part, yesteryear's behavioral objectives were not accompanied by assessment instruments specifically linked to those objectives. And one lesson I've learned for sure is that statements of intended instructional outcomes, whether called goals, objectives, or content standards, will have little impact on day-to-day instruction if not accompanied by relevant assessment devices. It is the presence of outcomes-based *assessments* that, because such assessments can determine whether an outcome has actually been achieved, will make a teacher attend to whether students master intended outcomes. Objectives without assessments are rarely more than rhetoric.

So, from an instructional perspective, that is, with respect to any impact on classroom instruction, behavioral objectives took a tumble because there were too many of them and they were not well measured. Unfortunately, those two drawbacks will almost certainly doom today's

content standards to a similar fate. The content standards currently in place, the cornerstones upon which many major reform efforts are being erected, are both *too numerous* and *poorly measured*.

Yet, because there has been so much positive energy invested in the creation of both national and state-specific content standards, is there any way to salvage this situation? I believe there is, and I think we need to get salvaging without delay. If we don't, content standards will be swimming alongside behavioral objectives in an Olympic-size pool of well-intentioned but feckless attempts to improve schooling.

Wishing Won't Make It So

First off, it is imperative for us all to recognize *clearly* what's going on when curricular specialists get together to isolate a set of content standards for a given subject area, usually at several different grade levels. In any subject area, there are all sorts of good things that kids can learn. But, of course, the educators who are chosen to isolate the content standards (knowledge and skills) needed by children will almost certainly be subject-matter specialists in the field for which the content standards are being developed. And almost all content specialists, be assured, *love* their content areas. Math specialists get tingly when Pythagorean Theorem is even mentioned. Some English teachers actually glow when explaining why a possessive pronoun is needed to modify a gerund.

As a consequence of content specialists' quite natural affection for their subject fields, most of the individuals who created today's collections of content standards end up *wishing students could learn it all*. And that, of course, is how we have ended up with wish-list content standards. It is understandable. But, from an instructional perspective, it's downright dumb. If we actually tried to get our students to master the laundry-lists of content standards set forth in various content areas during recent years, we would need much more instructional time to teach our students. There's way, way too much to teach in any set of wish-list content standards.

I do not fault the creators of wish-list content standards. Typically, they were not informed of the importance of assessing whether students' standard-mastery had occurred. And, oblivious of that important proviso, it is not surprising these content specialists ended up with as many content standards as they did.

The profusion of content standards, unfortunately, also creates the same sort of problem we encountered during the behavioral objectives era. There are so many content standards to assess that measurement folks really can *not* do a decent job in assessing students' mastery of them. If you don't believe me, just amble up to any seasoned test developers and hand them a set of 70 content standards in, say, science. Then ask them to devise a decent test to measure students' attainment of those standards *in 60 minutes of students' testing time*. You see, we do not have unlimited assessment time available to us. Given a finite period for assessment, the number of content standards to be assessed must be modest indeed. We must not allow assessment specialists to claim they can do a satisfactory job in measuring a multitudinous array of content standards. They can't.

Reality Recognition

We need to recognize the realities of our current situation. Most content standards contain collections of knowledge and skills (often called *benchmarks*) way too numerous to (1) allow for teachers' genuine instructional attention and (2) measure properly. And that criticism certainly includes those content standards that are thinly disguised attempts to collapse large sets of content standards under broader umbrellas. I have seen far too many instances where, for instance, a set of 100 content standards in a given field has been reformulated as 10 content standards, each of which (not surprisingly) embraces 10 distinctive sets of knowledge or skills. The original 100 content standards did not disappear at all. Such efforts to collapse too-lengthy sets of content standards are clear instances of *counterfeit coalescence*. The teacher must still pay attention to 100 different things, not 10 things. The teacher is no better off.

Another assessment reality is also important. Educators who organize their reform efforts around content standards need to determine *which* content standards have been mastered and *which* haven't. A measurement system that does not yield *per-standard or per-benchmark evidence of attainment* is, from an *instructional* perspective, meaningless measurement.

What this means, plain and simple, is that if there are too many content standards to be promoted—we can't have a per-standard measurement of them all. It is just not possible without an enormous increase in the time allowed for assessment, an increase that would take away from teachers' instructional time. Any pretense that a large-scale assessment program can provide satisfactory measurement of a plethora of content standards is just that—a pretense.

Salvation Time

I believe, however, there is a readily implementable salvation strategy we can apply to the assessment of wish-list content standards. Simply put, it requires us to *prioritize*. Here's how such a salvation strategy might work if you care to implement it.

First off, assemble a group of content knowledgeable educators to review an existing set of wish-list content standards. There may be advantages in having a brand new collection of folks perform this task rather than the individuals who created the original wish-list content standards. That's simply a practical judgment based on the well known proclivity of parents to find few faults in their progeny.

Next, announce that because the *per-standard assessment* of the existing set of content standards is patently impossible, it will be the task of these new reviewers to establish priorities among the standards. It is important to point out that *all* of the content standards may, indeed, be pursued *instructionally* by teachers. Moreover, the review group should be informed that a variety of *classroom assessment* techniques can be employed to measure students' mastery of any content standard not ultimately chosen for large-scale assessment. So, just because prioritizing the standards for assessment purposes is to take place, this does not mean that any particular content standard is doomed to disappear from teachers' instructional radar screens. Given sufficient energy and commitment, teachers can promote students' mastery of many more content standards than those measured by a large-scale assessment program.

Then have the reviewers sort all the content standards into three groups, those that are *desirable* for students to master, those that are *very desirable* for students to master, and those that are *essential* for students to master. Preliminary discussions of the meaning of each of these three ratings should precede reviewers' actual ratings. There are, incidentally, a number of procedural ways to rate a set of content standards. I prefer individual ratings followed by a modest amount of intermittent group discussion and an opportunity for reviewers then, if they wish, to alter their per-standard ratings. But many other approaches to standards-rating would work equally well. After the rating process is over, however, there should be three distinguishable sets of content standards, namely, those that the group has decided (possibly using simple voting procedures) were, for assessment purposes, *essential*, *very desirable*, and *desirable*.

Next, the reviewers should focus only on the *essential* content standards. Each reviewer should *rank*, from most important to least important, every essential content standard. As before, en route discussion among reviewers may or may not be permitted. These priority rankings should then be amalgamated so that, at the review-session's conclusion, a set of essential content standards exists—arranged from most to least important.

Ideally, the most highly ranked content standards would focus on truly significant—and teachable—skills such as a student's ability to write a powerful persuasive essay. Because only a limited number of content standards can be measured via a large-scale assessment program, it is imperative that those standards being measured deal with genuinely high-level cognitive skills. Although a "knowledge" may be required by students to enable them to carry out such high-level cognitive skills, it is difficult for me to think of a high-priority *essential* content standard that would focus exclusively on knowledge.

The reviewers may also be asked to rank-order the other two sets of content standards (that is, the desirable and very desirable ones), especially if sufficient time exists for that task. Such rankings can be especially helpful in guiding those who might undertake the creation of classroom assessments focused on *particular* content standards. Remember, if assessments are *not* standard-specific, then the result tends to be measurement mush—data of little utility for instructional purposes.

Finally, the prioritized essential content standards are turned over to the organization charged with the creation of a large-scale assessment. Often, this assessment will be of the high-stakes variety based either on (1) what happens to students because of their test scores or (2) whether public perceptions of educator effectiveness are based on the test's results.

The assessment group should then assess only those highest-priority content standards for which tests can be built to yield valid score-based inferences *per standard* given the test-administration time available for that purpose. What this means, of course, is that only a handful of the most important content standards will be assessed. But if the validity of educators' score-based inferences about students' standard-mastery is of any importance, then that's all a large-scale test can actually do. To contend that we can accurately assess a galaxy of content standards is assessment hypocrisy.

Also Needed: Teacher-Palatable Assessment Descriptions

Once the highest-priority essential content standards have been isolated, then there must be a meaningful collaboration between assessment and curricular personnel in order for them to produce an *assessment description* for each of the content standards to be assessed. These assessment descriptions must spell out, in teacher-palatable language, the essence of the cognitive demands on students required by any assessment built to measure the content standard involved. Indeed, the more variety there is in the kinds of assessment approaches used, the more generalizable must be the student's skill-mastery. And this means the teacher must aim more directly at promoting the student's *generalizable* skill-mastery. By isolating the essence of the cognitive demands associated with the skill embodied in a significant content standard, each assessment description presents a clear target of where the teacher needs to be heading instructionally. Such assessment targets, of course, do not stipulate *how* the teacher should get there. A century of research has shown us quite clearly that for different teachers, students, and instructional contexts, it is quite possible for dramatically divergent teaching styles to be successful.

What the creation of an assessment description for each assessed essential content standard does, of course, is provide teachers with the kind of clarity they need regarding the nature of the desired instructional outcomes. No teacher will find it necessary to teach toward the specific items on a test because the assessment description will make it clear what the cognitive skill is that the test's items are attempting to assess. Because only a modest number of high-import cognitive skills are being assessed by our high-stakes test, then we need to increase the likelihood that those skills will be well taught. The provision of teacher-palatable assessment descriptions, jointly crafted by curriculum and assessment personnel, will do just that.

What this approach leads to, of course, is more need for *efficient classroom assessment* of the many wish-list content standards not tested in a large-scale assessment program. Ideally, some central agency (e.g., state or district) will take the lead in providing teachers with a menu of standard-specific classroom assessments from which teachers could choose.

And, most importantly, teachers must be reminded that for purposes of instructional planning, *all* of a set of content standards—even of the wish-list genre—can be given instructional attention if their classroom time permits.

Or Else

If an approach akin to that suggested here for assessing wish-list content standards is not soon installed, it does not take a prophet to forecast what's apt to happen with standards-based educational reform. Nothing good.

Perhaps you will regard what is being proposed as some kind of minimalist abdication of our assessment responsibilities, an approach in which only a handful of super-important content standards are assessed via large-scale tests. Well, I want you to consider the implications of what I am recommending. It is an approach in which far more content standards will need to be

assessed via *classroom* tests because it is unequivocally clear that all wish-list content standards will *not* be assessed by a large-scale test. But, please believe me, they're not being assessed now!

Teachers will need to be selective, of course, in their choice of what content standards to measure with their classroom tests. But at least they can make those decisions in the clear recognition that the vast majority of wish-list content standards will, based on current assessment practices, never be measured in a way that teachers can do anything instructionally about them. Phony assessment tends to breed complacency. What I am proposing is an honest, instruction-focused approach to the assessment of today's content standards.

W. James Popham has, through the years, been an occasional contributor to the Bulletin commencing with an article in the Volume 45, December 1961 issue. An Emeritus Professor in the UCLA Graduate School of Education and Information Studies, Dr. Popham now lives on the island of Kauai in the state of Hawaii.