April 20, 2011

APPROPRIATE CURRICULAR AIMS FOR CONSORTIUM-CREATED ASSESSMENTS: FOUR KEY ATTRIBUTES

W. James Popham

University of California, Los Angeles

Our nation is currently caught up in the early stages of an enormous educational challenge to see whether the vast majority of our states can not only *adopt* identical curricular aims in mathematics and English/language arts (ELA), but can also devise *suitable instructional and assessment systems* linked to those aims.

This challenge arrived for us when, through the energetic initiatives of the Council of Chief State School Officers (CCSSO) and the National Governors' Association (NGO), a set of Common Core State Standards (CCSS) were recently produced that have been, or are currently being, adopted by all but a few states. The CCSS are curricular aims, that is, the CCSS identify the knowledge and skills (in ELA and mathematics) it is believed students should achieve as a result of their schooling. Currently, two assessment consortia, each composed of several dozen states, have received substantial federal funding to create assessments that can determine the degree to which students have mastered the curricular aims embodied in the CCSS. As I write this, those two assessment consortia are determining how they can best measure students' mastery of the curricular aims represented by the CCSS.

Because the assessments created by one or both of the consortia will most likely become the accountability tests routinely used in evaluating the success of most U.S. schools, the probable impact of these tests on American education is potentially enormous.

An Iceberg Awaiting its Titanic

It is my belief that if either or both consortia mistakenly set out to assess students' mastery of improperly conceptualized curricular aims, all their subsequent assessment-development efforts, despite zealous efforts and laudable intentions, will be certain to fail. Yes, choice of an unsound curricular conceptualization by either consortium will, of necessity, lead to the generation of assessments unlikely to improve our schools. Let

me illustrate how, in recent decades, we have seen such mistakes torpedo many promising school-reform efforts.

In almost every state, when curriculum specialists identify what they want their state's students to learn, those specialists come up with far too many targets. Indeed, most states' officially approved curricular aims tend to resemble "wish lists" rather than realistic collections of what students can actually learn while in school. The curricular specialists who author those lists of curricular aims typically identify all the skills and knowledge that they *wish* their state's students will learn in school. But the resultant numbers of "wished-for" curricular aims invariably turn out to be too many to successfully teach (at any depth) during the available instructional time or to appropriately assess during the available assessment time. Putting it differently, teachers can't teach so much stuff meaningfully in the instructional time they have available to them. Moreover, measurement folks can't assess students' mastery of so much stuff validly in the amount of assessment time available to them. I have bemoaned elsewhere (Popham, 2009) the negative results when educators' curricular eyes are bigger than their instructional and/or assessment stomachs.

The architects of the CCSSO and NGA curricular aims had decided, with much applause from most bystanders, to create curricular targets that were "fewer, clearer, and higher." Yet, while creators of the CCSS have done reasonably well on the "clearer" and "higher" dimensions, they appear to have scored below-basic when it came to "fewer." Put candidly, there are simply too many curricular aims—represented at varying levels of generality—now contained in the CCSS. Thus, if either consortium tries to devise assessment approaches intended to measure *all* the curricular aims embodied in the CCSS, this will constitute a blunder from which the erring consortium will never recover.

One significant function of curricular aims is to help teachers direct their instructional efforts, and to help students direct their learning efforts, toward those aims. If too many curricular aims exist, such a profusion of goals will most certainly overwhelm both teachers and students. By the same token, because one of the substantial payoffs of properly devised assessments is to supply teachers, students, and students' parents with accurate information about students' mastery of curricular aims, when assessment specialists are presented with too many curricular aims to be measured, they have little recourse but to assess those too-numerous aims superficially rather than in a manner from which it is possible to arrive at accurate inferences about students' per-aim mastery. Thus, attempting to construct assessment systems based on the entire collection of CCSS aims constitute an irreversible error.

Instead, what both consortia's leaders should do is base their assessment-development strategies on a collection of curricular aims that, though *directly derivative* from the CCSS, have been reconceptualized in such a manner that those aims will not only stimulate improved classroom instruction, but can also spur the creation of assessments capable of better informing educators (and the public) regarding how effectively our students have been taught.

Four Attributes of a Set of Appropriate Curricular Aims

But what would a set of assessment-appropriate curricular aims look like? In the remainder of this brief analysis, I will identify four qualities of a collection of curricular aims that could serve as suitable assessment targets for either assessment consortium. Each of these four attributes could be dealt with in greater detail than will be found here, complete with illustrative curricular aims that would either adhere to or violate each attribute. However, in this short commentary, the four attributes will only be identified and tersely described. Let's turn, then, to the first of these four attributes.

Attribute 1: Numerical Manageability

Only a manageable number of curricular aims should be measured by any assessment destined for summative use.

As implied earlier, almost a generation's worth of state-built accountability tests have functioned ineffectively because they set out to assess students' mastery of an excessive number of curricular targets. When teachers are presented with too many curricular aims, they are faced with an insoluble instructional dilemma, that is, they must either give superficial attention to all such aims or must give no attention to some. Typically, because the accountability tests built to assess students' mastery of those too-numerous aims will only *sample* from the complete array of assessment-eligible curricular aims, much effort is often expended by teachers in trying to *guess* which curricular aims will be measured on an upcoming accountability test.

But from an instructional perspective, the problem with too many curricular aims is even more insidious. Teachers—and students too—can benefit from having clearly in mind an idea of what should be learned by the end of a semester or a school year. Given a reasonable number of curricular targets between, say, six-to-ten, most teachers (and most students) can keep this modest collection of curricular targets in mind. A fourth-grade teacher, for example, who wants her fourth-graders to master eight important mathematics curricular aims and seven significant ELA aims can use this intellectually manageable number of curricular targets as a guiding curricular framework for an entire

school year. A teacher can make sense of, accurately comprehend, and keep track of ten or so curricular targets. Few teachers can make sense of, accurately comprehend, and keep track of four or five *dozen* curricular targets.

But if every curricular target currently embodied in the CCSS is not assessed, won't this mean those non-assessed curricular targets will therefore receive no instructional attention at all from the nation's teachers? This is a worrisome question, and the answer to it is *not necessarily.* We all know that teachers will tend to emphasize instructionally those curricular targets that have been designated for assessment by high-stakes accountability tests. But this does not mean that teachers give *no attention whatsoever* to curricular content not slated for subsequent high-stakes assessment. Accordingly, if a truly *manageable* number of to-be-assessed curricular targets were being pursued, is it not likely that teachers would be able to promote students' mastery of those aims *more efficiently* and, thus, would have at least some instructional time available to promote students' mastery of many non-assessed curricular goals?

The argument that, when confronted by a modest number of significant to-be-assessed curricular aims, teachers will never deal with non-assessed curricular goals is every bit as specious as contending that if a huge array of curricular aims has been designated as assessment-eligible, then surely those curricular aims will all be successfully promoted as a consequence of this designation .

Another substantial payoff from focusing assessments on some, not all, CCSS aims is that it will then be possible to include a sufficient number of items on any assessments being used summatively to provide teachers, students, and students' parents with more meaningfully accurate estimates of the degree to which *specific* curricular aims have been mastered by each student. By assessing a student's per-aim mastery with 6-10 items, we can often get a reasonable fix on a student's master of each assessed aim. While it is true that the measurement of a student's mastery of a key skill might not be *definitively* demonstrated by, say, a half-dozen items dealing precisely with that skill, such an inference regarding the student's skill-mastery is certain to be more valid than would an inference based on only one or two items. With fewer curricular aims to assess, it will be usually be possible to do a decent job of measuring students' per-aim achievement levels.

The key question associated with this first attribute of is simply: "How many curricular aims are, in fact, *unmanageable*?" To answer this question, there is no magic formula that, if properly employed in diverse content areas, will dutifully spit out a "too-many aims" number. Rather, the answer boils down to a circumspect judgment regarding the number of curricular aims that can realistically be taught by teachers in the instructional time that's available to them. And the "teachers" to in the preceding sentence refer to

solid, professionally competent teachers—not super-star siblings of Socrates. In essence, a manageable number of curricular aims should reflect the skills and knowledge that capable teachers can get the vast majority of students to master at a meaningful, not superficial level, during a school year.

So, for both instructional and assessment reasons, this first attribute of a collection of assessment-appropriate curricular aims is, in truth, a *sine qua non.* Without numerical manageability, a collection of curricular aims simply cannot serve as the foci for a truly successful consortium-created assessment system.

Attribute 2: Educational Import

Consonant with the instructional time available, only the most important curricular aims should be designated for assessment.

If one accepts the reasonableness of focusing a summative-assessment system on only a modest number of curricular aims, it then becomes imperative to make sure those aims measure the most important things that students of a given age should know and be able to do. In other words, because there will be fewer—perhaps far fewer—curricular targets designated as eligible-to-be-assessed targets, then those fewer, assessment-anointed curricular goals must be extraordinarily important.

To illustrate, in ELA we surely want students to be able to generate, from scratch, original age-appropriate compositions, and to do so with increasing sophistication as they grow older. In mathematics, not only do we want students to be able to perform the four basic arithmetic operations properly, but to be capable of employing those operations in the solution of previously unencountered problems. In other words, when we have only a limited number of curricular targets to adopt, we must be super-sure that those targets are the most significant—the most life-influencing—that we can identify for our students.

Because of the need to carefully isolate the outcomes that can realistically be accomplished at each grade level in mathematics and ELA, any set of potential to-be-assessed curricular aims—all of which should be directly drawn from the CCSS—must be rigorously *prioritized* not only by ELA and mathematics subject matter experts, but also by experienced teachers who are in a position to judge what can actually be accomplished instructionally with students of a given age during the available instructional time.

Because, as the CCSS aims are currently formulated, they often do not seem suitable for a less-is-more isolation of assessment-appropriate curricular aims, it will often be necessary not only to prioritize certain of the CCSS targets but, in a number of instances, to coalesce certain of those targets into more comprehensive curricular goals that subsume lesser goals. (See Attribute 4 below regarding how such subsuming might be accomplished.)

This second attribute of a set of assessment-appropriate curricular aims, a set that would be suitable for the development efforts soon to be undertaken by both consortia, involves approaching the curricular preferences embodied by the architects of the CCSS, recognizing the nature of their curricular aspirations, then reflecting those preferences as faithfully as possible in assessments that will contribute to both enhanced instruction and more accurate evaluations. If a consortium's leaders subscribe to Attribute 1 (numerical manageability), then this second attribute of educational import should be regarded as a required corollary of that first attribute.

Attribute 3: Instructional Amenability

All curricular aims chosen for summative-function assessments must be addressable instructionally.

Some things we do in life are more attributable to the smarts we inherit from our parents than from what we have learned in school or elsewhere. These inherited qualities are usually cognitive skills that, because of the gene-pool lottery, some students possess more of than do other students. Certain students, let's face it, are born with bigger slices of the academic-aptitude pie than are other students.

This third attribute of assessment-appropriate curricular aims hinges on whether students can, in fact, *be taught* to accomplish what's represented in an under-consideration curricular aim. Care must always be taken to avoid the inclusion of curricular targets whose mastery, in a very direct fashion, hinges most heavily on students' innate smartness. Curricular aims awash in inherited intelligence should surely be eschewed.

What this means, in practical terms, is that those who are considering a curricular aim as a potential target for assessment must give at least some thought to the following question: "How can students be taught to master this aim?" If no clear answers to this question come to mind, then a re-think regarding the potential curricular aim's suitability seems warranted.

Attribute 4: Instructional Separability but Unitary Assessability

A curricular aim should be unitarily assessable but, if it incorporates subcomponents, each of those should be instructionally separable.

This final attribute of an assessment-appropriate curricular aim, at least an aim that's under consideration as a target for summative-function assessment, is particularly important but somewhat difficult to explain. Perhaps an early-on illustration of what's being advocated will help.

When we assess students' skill in composing original essays of various sorts, we often try to gauge a student's composition prowess by asking the student to generate a writing sample that we can then evaluate not only *holistically*, but also *analytically*. That is, we can arrive at an overall judgment (holistically) about a given student's composition skill, but can also reach separate judgments (analytically) regarding such evaluative factors as the student's organizational skill, mastery of written mechanics, and knowledge of the composition's content. Whereas, based on a student's writing sample, a *unitary* judgment can be made regarding the student's overall composition skill, this overall judgment hinges on a composite appraisal of distinct dimensions, each of which can be *separately* addressed instructionally. A curricular aim dealing with students' mastery of an important skill, such as being able to create an original essay, exemplifies this fourth attribute, that is, it is a curricular aim that is instructionally separable but unitarily assessable.

The reason this fourth attribute is so very important hinges, as usual, on the instructional consequences of the attribute. For assessments to have a positive impact on students' educations, those assessments must focus on a set of curricular aims that teachers can not only readily understand, but can vigorously promote instructionally. By satisfying the first three attributes identified here, we can be sure that our to-be-assessed curricular aims are not too numerous, deal with truly important outcomes, and can be tackled instructionally. But these positive features evaporate if the modest number of curricular aims being pursued is, in fact, not modest at all. What if a particular curricular aim is, in reality, not one homogeneous aim but, instead, is pretending to be a unitary curricular aim when it actually embraces several quite different student outcomes. This situation is often encountered when curriculum architects rely on "strands" to describe the nature of a set of related, but dissimilar cognitive skills.

For instance, suppose a mathematics curricular aim calls for students (as its Subcomponent 1) to be able to arrive at reasonable estimates regarding the area contained in certain geometric shapes for which *insufficient* area-calculation data are

given, but also asks students (as its Subcomponent 2) to be able to calculate areas of geometric shapes when all the necessary data *are* on hand. The subskill embodied in Subcomponent 2 is related to, but substantively different from, the Subcomponent 1 estimation subskill. Yet, we might find some curriculum specialists who try to lump both of these subskills under a strand-heading such as "Geometric Areas." Although each of these two subskills could be tackled instructionally, students' mastery of the two subskills could not be assessed at one time with a unitary assessment procedure. It would be possible, of course, to create a single test containing two conceptually distinct subtests, one for each subskill, but this would not be *unitary* assessment. And this is why, with some justification, it has been said that curricular "strands" usually mask, not clarify the nature of what we expect our instructed students to accomplish.

We definitely want any subcomponents of a curricular aim to be instructionally addressable, for if it turns out that students are having difficulty mastering a particular curricular aim, then we can follow up with diagnostic tests to help us discern which subskills or bodies of enabling knowledge appear to be causing the problem. Once diagnosed, then teachers can zoom in instructionally to deal with those areas of weakness. And earlier, when a teacher's instructional plans are being initially developed, it is advantageous if the subcomponents of a more comprehensive curricular aim can be addressed directly. Ideally, of course, if a curricular aim embodies subcomponents such as subskills or bodies of enabling knowledge, there should not be too many of those subcomponents. With respect to a curricular aim's subcomponents, just as was true with curricular aims themselves, less typically turns out to be more. But no matter how many subcomponents a given curricular aim contains, it is crucial to for students' collective mastery of those subcomponents to be holistically assessable—all at one time—yet for each subcomponent to be instructionally addressable.

Only One Bite of this Apple

If one were to stake out a continuum of high-stakes assessment, it is impossible to imagine a higher-stakes set of educational tests than the assessments soon to be created by the two assessment consortia. If done well, these new tests could become catalysts truly capable of fostering a new era of more effective, assessment-abetted schooling in the United States. If done badly, the new tests will have wasted an enormous amount of money and, worse, will have deflected America's educators from pursuing other improvement strategies that might have bettered our schools.

But be assured, the two consortia will have *only one shot* at success or failure. Too much is at issue here for national policymakers to allow the unsuccessful consortia a do-over. Accordingly, the selection of the curricular aims to be measured by consortia-

created assessments is altogether imperative. The wrong curricular choices by a consortium's leaders will definitely cripple that consortium's measurement mission.

Each of the curricular aims chosen as assessment targets must satisfy *all four* of the attributes identified here. Failure to do so will result in the creation of ineffectual assessments, incapable of accomplishing the good we all hope these tests will promote. However, this wonderful opportunity will come our way but once. We must not muff it.

Reference

Popham, W. James, 2009, *Unlearned Lessons: Six Stumbling Blocks to Our Schools' Success*, Cambridge, MA: Harvard Education Press.