MEASUREMENT INCORPORATED

EXCELLENCE IN ASSESSMENT™

Aligning Curriculum, Instruction, and Assessment

Michael B. Bunch, Measurement Incorporated

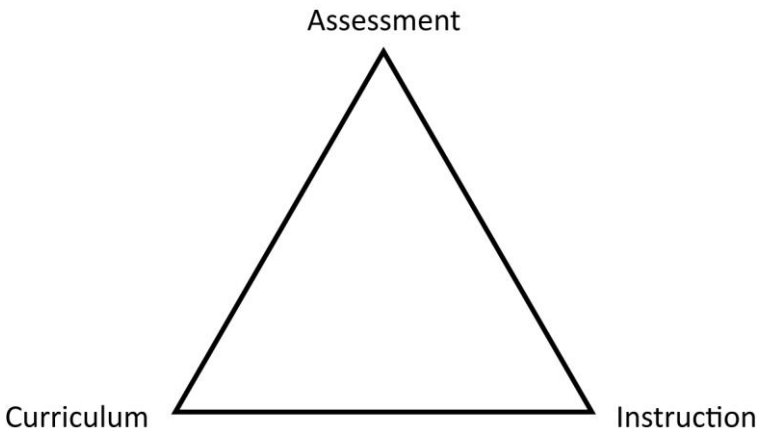
A key component of educational achievement test validation is alignment of the test to both curriculum and instruction. By alignment, we mean the degree to which the items of the test, both individually and collectively, match the structure and intent of the curriculum and instruction. This paper has several purposes:

1. To provide an overview of the goal and process of alignment
2. To raise some important questions we need to answer as we create tests to assess college and career readiness, and
3. To expand our view of alignment of tests of college and career readiness

Goal and Process

The goal of alignment is to make curriculum, instruction, and assessment work toward the same ends. Generally, we start with curriculum, lay out goals for instruction, instruct to achieve those goals, and assess to determine how successful we've been in achieving the goals set forth in the curriculum. We tend to think of the process in terms of a triangle, as shown in Figure 1.

Figure 1: The Alignment Triangle



The three components of Figure 1 are interactive; any of the three can and should inform the other two. For example, in addition to curriculum driving instruction and assessment, it is also possible that instruction can provide feedback for improving curriculum and refining the format of test items. Similarly, assessment can identify weaknesses in instruction that can be corrected and detect holes in curriculum that can be filled.

© 2012 Measurement Inc

Clarifying Terms

For the purposes of this paper, I confine my remarks to summative assessment and offer the following working definitions.

- **Curriculum**: The written set of educational outcomes and associated content that students are to learn. This will include the knowledge, skills, and abilities we expect students to acquire or master after a period of appropriate instruction. Curriculum may include prescribed activities and procedures designed to bring about mastery of that content.

- **Instruction**: What actually happens at the classroom level (and at home and in the community)? This includes all the activities of teachers, students, aides, parents, and others involved in transmitting to students a set of knowledge, skills, and abilities. For formal evaluation purposes, however, only those activities under the control of educational authorities will be considered.

- **Assessment**: The formal process of gathering, analyzing, and reporting standardized information about the acquisition of knowledge, skills, and abilities for a group of students. For the purposes of this paper, we will focus on summative assessment (i.e., formal assessment at the end of a prescribed period of instruction, such as a semester or school year).

The Process

Alignment as a formal process with quantifiable outcomes is fairly new. For decades, test developers prepared blueprints and specifications, wrote items, assembled tests, and simply compared the final versions to the original blueprints. Quantification was typically limited to comparing percentages of items in cells of a final test blueprint with those of an ideal form. At the item level, test developers submitted items to committees of reviewers to make sure each item conformed not only to the content of the curriculum but to the format and process of the instruction as well.

In the past decade, Andy Porter and Norman Webb have contributed significantly to the quantification of the alignment process. Andy Porter introduced the Survey of Enacted Curriculum (SEC, seconline.wceruw.org). Norman Webb has given us the Webb Alignment Tool (WAT, wat.wceruw.org). Both employ Webb's depth of knowledge (DOK, dese.mo.gov/divimprove/sia/msip/DOK_Chart.pdf) scale. With these tools, educators are able to plot curriculum, instruction, and assessment on a two-dimensional grid to create a variety of useful visual displays. Particularly useful is the side-by-side comparison of a map for a curriculum and its associated test. The peaks and valleys of one map should match those of the other. To the extent that they do, the test is aligned to the curriculum. Similarly, the map of a local curriculum or even a segment of instruction can be compared to a state curriculum map.

© 2012 Measurement Inc

The mapping procedure for both the Porter and Webb approaches involves groups of educators evaluating and indexing large amounts of material. As with any procedure that involves human judges, the final product typically represents general consensus rather than perfect agreement. Fortunately, the intra-group differences are generally small enough that the final product is quite a valid and reliable statement about the curriculum, instruction, or assessment.

Some Important Questions

Since the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (SBAC) began developing plans for the assessments of 2014–15, the alignment conversation has focused on the tests and the Common Core State Standards (CCSS). So far, we haven't heard much about instruction, other than that it is assumed that states, districts, and schools will implement these standards. It is an assumption that bears testing.

"Will students have an opportunity to learn?"

- **Are the CCSS being implemented, and if so, how?**

In a recent white paper, *Cut Scores for 21st Century Assessment*, I mentioned that states are implementing the Common Core State Standards in different ways and on different schedules. If we develop tests well aligned to the CCSS but fail to provide instruction equally well aligned, the results of the tests become meaningless because the triangle is missing one of its points.

- **Are the ideals of the test designers reflected in instruction?**

Both PARCC and SBAC assessments are based on an evidence-centered design. Moreover, both consortium assessment-development plans assume instruction based on cognitive research. Indeed, such an assumption is well warranted, as the No Child Left Behind bill refers to "scientifically based research" over 100 times. But is this assumption warranted in terms of what teachers are actually teaching? Has anyone told them that their students will be assessed in this manner and that they should be teaching in a way that reflects best practice as defined by cognitive research? At this point, it is too late to expect teacher-training institutions to assume this responsibility. It is up to states and districts to make sure teachers—and ultimately students—are ready. Will they be?

- **How will test design be informed by current instruction?**

Both consortia have built into their development schedules reviews of items by in-service educators. That is commendable. The important question now is, "How will we incorporate the experiences of those educators into the refinement of the items and tests?" This question has implications not only for how we record and utilize their ratings of items but how we design the rating forms, conduct the training, and manage the review meetings. The following exchange is synthesized from several item-review meetings we have conducted over the past 30 years. It illustrates what can and does happen at these meetings:

© 2012 Measurement Inc

Chair: So we are calling item 12 DNU (Do Not Use). And our reason?

Teacher 1: Does not align.

Chair: OK, does not align. And the comment we will put beside that entry?

Teacher 2: We don't actually teach this in fifth grade. It's more a sixth grade objective.

Chair: Can we say that? It's on the state curriculum for fifth grade.

Teacher 3: Yes, that's true. A friend of mine was on the committee that prepared the standards, and she tried to get them to move it, but they wouldn't listen.

Teacher 2: And besides, the format is different from the way we teach it.

Chair: So we mark item 12 DNU and note that it is incorrectly formatted and off-grade.

All: Yes, that sounds good.

No doubt, the review sessions for PARCC and SBAC will not go quite like this, but it is important to begin thinking now about the training, forms, and recording procedures of the review sessions in order to make sure the tests that emerge from those sessions are aligned to the curriculum and instruction or that any misalignment between curriculum and instruction can be corrected before the spring of 2015.

- **Will students have an opportunity to learn?**

The United States Court of Appeals (Fifth Circuit) ruled on May 4, 1981, that "The State may not deprive its high school seniors of the economic and educational benefits of a high school diploma until it has demonstrated that the SSATII (the Florida minimum competency test) is a fair test of that which is taught in its classrooms." (*Debra P. v. Turlington* 474 F. Supp. 244 (M.D. Fla., 1981)). Even though the PARCC and SBAC tests may not carry diploma sanctions, the court's decision set clear ground rules for test development when postsecondary opportunities are at stake:

© 2012 Measurement Inc

1. Students must be told the specific objectives on the test.
2. Students must be given instruction in these objectives.
3. This instruction must be rational, orderly, and cumulative.
4. Students must be given time to master these objectives.
5. Instruction must include accountability.
6. Students must be afforded opportunity for remedial instruction.

The *Debra P*. decision resulted in a two-year delay in the implementation of the tests, during which time Florida was required to conduct a series of surveys to satisfy the court that students had indeed had an opportunity to learn. As the CCSS are rolled out, who is monitoring their implementation at the classroom level? Who will be able to state unequivocally that all students have had an opportunity to learn the assessed content by the spring of 2015?

We have a lot to think about in the next couple of years: Evidence-centered design, instruction based on scientific research, making sure everyone gets a chance to learn all this stuff, and making sure we test what was taught. Could our plates possibly get any fuller? Yes.

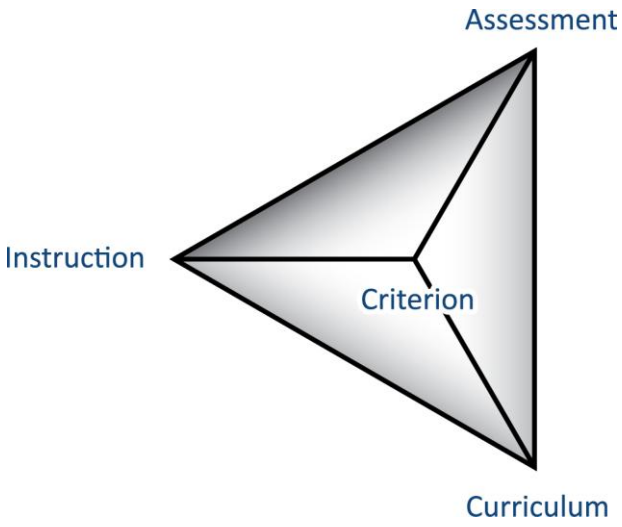
Expanding Our View

And now for something completely different…

To this point we have viewed alignment in two-dimensional space (i.e., a triangle whose three points are curriculum, instruction, and assessment). With the advent of college and career readiness as our ultimate goal, we have for the first time in a long time a viable external criterion by which all three points can be validated. The introduction of this fourth point moves us from two dimensions to three and transforms our alignment triangle into an alignment pyramid, as shown in Figure 2.

© 2012 Measurement Inc

Figure 2: The Alignment Pyramid



For most of the past century, college readiness has been defined by the contents of the College Board Scholastic Aptitude Test (now Scholastic Assessment Test, but still SAT) and the American College Test (ACT) from American College Testing. Both test batteries were created in accordance with content blueprints that emphasized verbal and quantitative abilities, but they were refined over time by validation against first-year college grades. Employment tests followed a similar pattern of development and validation, starting with job task analysis and ending with validation against supervisor ratings or other quantifiable criteria such as sales commissions or acceptable units produced per hour.

We have assumed up until now that the CCSS embody all relevant aspects of college and career readiness. As the process of test development unfolds and as other voices enter the conversation, we may wish to test this assumption from time to time. David Conley and others have pointed out several aspects of college readiness that are not explicitly addressed in the CCSS. Career readiness is also open to further discussion.

If we decide that validation against external criteria is indeed appropriate, how shall we go about it? Fortunately, the answer has been around for quite some time. For PARCC and SBAC tests, we would do well to follow the examples set by the College Board, ACT, and the National Longitudinal Survey (NLS) over the past several decades:

- Follow graduates who have taken these tests as they enter college and the workplace.
- Obtain first-year grades, employer evaluations, and other performance information.
- Validate high school test scores against these external criteria.

© 2012 Measurement Inc

In short, we find out exactly how college and career ready these students are, not just in terms of scores on a test, but in terms of real-life experiences beyond high school. To validate scores for students in the lower grades, we compare their scores on PARCC and SBAC tests to later performance in high school. Thus, in effect, we develop a set of pathways from third-grade reading and mathematics achievement to adult accomplishment, in a series of small steps over a very long period of time.

In this protracted process, we modify the tests as necessary to make them more criterion valid as well as more construct valid. If we change the tests that we know are aligned with curriculum and instruction, then it may also be necessary to modify those two points on the pyramid so that all are aligned with adult success.

There is another aspect of the validation process historically employed by college and career assessments that bears mentioning here. From their beginnings, the College Board and ACT have encouraged each postsecondary institution to conduct its own validation study in order to make sure the candidates it admits fit that institution's goals, curriculum, and instruction. Similarly, the American Psychological Association encourages each employer to conduct its own validation studies. The net result is that each institution or employer uses information from the tests in a slightly different way, such as by assigning different weights to the various tests and subtests or combining test scores with other indicators such as grades, noncognitive measures, or personal attributes.

"In short, we find out exactly how college and career ready these students are, not just in terms of scores on a test, but in terms of real-life experiences beyond high school."

What this suggests for PARCC and SBAC tests is that in 2015, the job will not be over; instead, it will have just begun. Postsecondary institutions, employers, our military, and others concerned with the readiness of young people exiting high school will begin conducting their own validity studies. Districts will develop local models linking elementary to middle to high school performance. Over time, a body of knowledge will accumulate that will allow PARCC and SBAC or their successors to refine tests at all grade levels. At the same time, states, districts, and schools will have an opportunity to revise curriculum and instruction as necessary to keep up with the world beyond high school. The Common Core State Standards will be a living document attuned to that world and evolving with it.

© 2012 Measurement Inc

Suggestions for Further Reading

In addition to the information referenced in this paper, the following print materials will also be extremely helpful.

Debra P. v. Turlington 474 F. Supp. 244 (M.D. F1a.,1981). This landmark case has set the agenda for every test-development effort since 1981. It is the foundation for opportunity-to-learn surveys and even affects the design and conduct of item-review sessions. Anyone who develops high-stakes tests should be familiar with this important legal precedent.

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. [Presidential Address] *Educational Researcher*, 31 (7), 15–21. This is one of the first widely-available expositions of the Porter approach to alignment, complete with tables and graphs illustrating his approach. It includes several maps based on his alignment software.

Porter, A. C., Polikoff, M. S., Zeidner, T., & Smithson, J. (2008). The quality of content analyses of state student achievement tests and content standards. *Educational Measurement: Issues and Practice*, 27 (4), 2–14. This article provides an up-to-date view of the Survey of Enacted Curriculum developed by Porter and his associates and provides examples and two illustrative appendices.

© 2012 Measurement Inc