

## Essential Information for Education Policy

# Research Points

## Standards and Tests: Keeping Them Aligned

The “No Child Left Behind” Act requires that states align K–12 assessments with their academic standards for what students should know and be able to do. States and school districts now face a test of their own: What is the best way to evaluate and to improve test alignment?

In the early 1990s, studies<sup>1</sup> found that students’ gains on high-stakes tests often were not reflected on different assessments that measured similar content. Apparently, teachers were narrowing their instruction to focus on the specific items in high-stakes exams, instead of on the broader knowledge and skills that the tests were designed to measure.

More recent studies by the RAND Corporation<sup>2</sup> and the National Board on Educational Testing and Public Policy<sup>3</sup> show that teachers reallocate their time to emphasize the subjects on state tests at the expense of nontested subjects. Even within content areas, teachers shift the focus of lessons to stress the material on the state exams. This is especially problematic when a single form of a test is used repeatedly.

Because tests can affect what is taught so powerfully, policymakers need to ensure that state tests actually measure what they were intended to measure. Otherwise, educators will be chasing the wrong goal, and policymakers will not have a good gauge of whether schools are producing the desired results. That is the essential idea behind “alignment.”

Alignment is at the core of standards-based education. The theory is that student achievement will improve if all parts of the system pull in the same direction. Under such a system, curriculum materials, teacher professional development, and classroom instruction all reflect a state’s standards. Tests designed to measure students’ performance of those standards clearly signal what is expected and probe whether students are learning the desired content and skills.

Welcome to the  
inaugural issue of  
*Research Points!*

We look forward to  
receiving your comments at  
[ResearchPoints@aera.net](mailto:ResearchPoints@aera.net).

Today's calls for alignment are built upon a foundation of more than 70 years of research<sup>4</sup> on the development, evaluation, and use of tests. *Standards for Educational and Psychological Testing*,<sup>5</sup> the recognized authority on educational testing, stresses that a "valid" test must show that it actually measures the constructs — knowledge, skills, abilities, processes, and characteristics — it was intended to measure. When a test is used to measure the achievement of curriculum standards, it is essential to evaluate and document both the relevance of a test to the standards and the extent to which it represents those standards.

Studies of alignment measure the match, or the quality of the relationship, between a state's standards and its tests. That match can be improved by changing the standards, the tests, or both.

## How To Know if a Test Is Aligned

Recent studies of alignment between state tests and standards demonstrate that alignment can be measured reliably by using agreed-upon criteria and procedures.

While their methods may differ, generally the studies ask these questions about alignment:

- ▶ **Does the test's content match the content (topics and skills) in the standards?** In other words, each test item should correspond to an objective in the standards. Similarly, key ideas in the standards should appear on the tests.
- ▶ **Do the tests and standards cover a comparable "range" or breadth of knowledge, and is there an appropriate "balance" of knowledge across the standards?** Alignment studies look at whether a test fairly and effectively samples across the range of objectives described in a state's standards instead of focusing on only a few objectives or disproportionately sampling students' knowledge of some objectives but not others.
- ▶ **Does the level of cognitive demand or challenge called for in the standards match that required for students to do well on the assessment?** For example, if the standards require students to synthesize information and explain their thinking, but the test items only

ask students to recall facts, the standards and tests would not be well aligned.

- ▶ **Does the test avoid adding material that is irrelevant to the standard supposedly being assessed?** For example, a test item may have an inappropriate "source of challenge," requiring a student to read and understand a long passage about space travel, when it is seeking to measure a student's knowledge of how to estimate distances and travel times.

## How Much Alignment Is Enough?

While there is no easy answer to this question, recent studies have yielded a remarkably consistent set of findings about the need to improve standards-to-test alignment.

According to researchers, many content standards are too vague to determine whether a particular test item corresponds to the standard or not. Andrew C. Porter of the University of Wisconsin-Madison, for example, found that states with separate standards for each grade generally had tighter alignment than states whose standards covered a range of grades, such as 3–5, 6–8, and 9–12.<sup>6</sup> In some instances, state standards may not be sufficiently specific to allow an assessment to be aligned tightly with them or to provide adequate guidance for teachers.

But it is possible to go too far in the direction of specificity. Some states list so many standards that tests cannot measure them all. Both Norman L. Webb of the University of Wisconsin-Madison<sup>7</sup> and Achieve, Inc.,<sup>8</sup> a Washington-based nonprofit organization, have found that it is harder for tests to reflect the full range of knowledge included in state standards if content expectations are spread across a large number of standards.

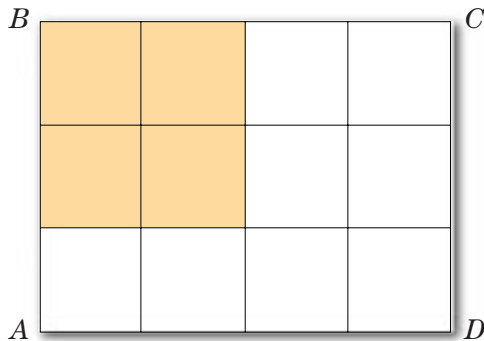
Studies also have found that many tests fail to measure adequately the cognitive complexity or "depth of knowledge" described in state standards. For example, in an analysis of 8th-grade math standards and tests in one state, Porter found that the test questions concentrated on length and perimeter, area and volume, and the performance of routine procedures. But the tests generally neglected more complex topics and cognitive demands included in the standards, such as angles or solving nonroutine problems.

*continued on page 4*

# Common Challenges in Test Alignment

## Small Changes in Test Items Can Shift Cognitive Demand

This item is intended to measure a student's ability to use two-dimensional region models to describe numbers. The level of thinking required is very different for two similar items.



*What fraction of ABCD is shaded?*

Requires less thinking because it explicitly defines the square units to make the shaded area easier to measure.



*What fraction of ABCD is shaded?*

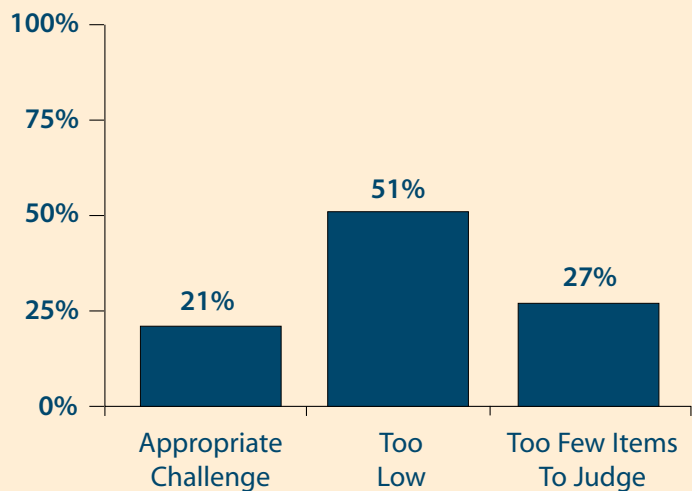
Increases the challenge by requiring that the student make measurements to determine what portion of the figure is shaded.

### Noteworthy

Many assessments focus disproportionately on simpler standards. A five-state review of English and math standards and tests by Achieve, Inc., concluded "The most challenging standards and objectives are the ones that are undersampled or omitted entirely ... [and those] that call for high-level reasoning are often omitted in favor of much simpler cognitive processes."

*Figures adapted from Rothman, et al.<sup>8</sup>*

### Tests' Levels of Challenge Often Do Not Match Standards — Results from a Typical State



### Facts at a Glance

- ▶ Teachers direct their lessons to what appears on tests. If tests are not well-aligned to standards, the standards may receive little attention.
- ▶ Many of today's tests are aligned poorly. The most cognitively demanding standards are not well measured.

Similarly, in an analysis of math standards in three states, Webb found that nearly all of the states failed to measure adequately the “depth of knowledge” described in their standards.

“What is included and excluded is systematic,” Achieve determined, based on its five-state review of English and math standards and tests. “The most challenging standards and objectives are the ones that are undersampled or omitted entirely ... [and those] that call for high-level reasoning are often omitted in favor of much simpler cognitive processes.”

As teachers try to match their instruction to state tests, the study warned, there might be an increased focus on the least demanding aspects of a state’s standards and a decreased focus on the high demands that define a rigorous curriculum.

## Concluding Point

While specific findings may vary from study to study, all of the research points to one central conclusion: Alignment needs to be improved. In some extreme cases, studies have found that alignment between state standards and tests is

### Bibliography

- 1) For a review of the studies, see Linn, R.L., “Assessments and Accountability”, *Educational Researcher*, Vol. 29, No. 2, pp. 4–15.
- 2) Stecher, B.M. and Barron, S., “Unintended Consequences of Test-Based Accountability When Testing in ‘Milepost’ Grades,” *Educational Assessment*, Vol. 7, No. 4, 2001, pp. 259–282.
- 3) Pedulla, J., et al. (2003). *Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers*. Boston: Boston College, The National Board on Educational Testing and Public Policy, Lynch School of Education.
- 4) Cronbach, L.J. (1989). *Essentials of Psychological Testing* (5th ed.). New York: HarperCollins.
- 5) American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- 6) Porter, A.C. (2002). “Measuring the Content of Instruction: Uses in Research and Practice,” *Educational Researcher*, Vol. 31, No. 7, pp. 3–14.
- 7) Webb, N.L. (1999). *Alignment of Science and Mathematics Standards and Assessments in Four States*. (Research Monograph No. 18.) Madison, WI: University of Wisconsin-Madison, National Institute for Science Education.
- 8) Rothman, R., Slattery, J.B., Vranek, J.L., and Resnick, L.B. (2002). *Benchmarking and Alignment of Standards and Testing*. (CSE Technical Report 566.) Los Angeles: University of California-Los Angeles, Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing.

**First**, states need to think about alignment from the start. Even before developing a test, states need to ask whether they have too many standards, whether the standards are clear enough, and which objectives are most important and should be emphasized. States can use the criteria in the alignment studies to help build test blueprints and to create better assessments from the beginning.

**Second**, state officials should not just rely on test developers to tell them whether their tests are aligned adequately with state content standards. It is essential to conduct an analysis that uses multiple criteria and agreed-upon procedures, such as those described in this brief. In addition to the work of Achieve, Webb, and Porter, alignment studies have been conducted by organizations including the Buros Institute; the Center for Research on Evaluation, Standards, and Student Testing; and several federally financed regional laboratories.

**Third**, as policymakers examine alignment between state standards and tests, they should consider whether the standards themselves need revision: Are they challenging enough, specific enough, and limited enough in number to sufficiently guide teachers and test developers?

**Fourth**, states should update their alignment studies whenever the standards or the tests change. And they should commit to using multiple forms or versions of an assessment, not just a single form. These measures make it harder to “teach to the test” and increase the likelihood that the test will reflect the standards adequately.

so weak that the standards from one state more closely match the tests used in another state.

Of course, no test can reflect fully a state’s expectations for its students. Some standards are difficult to measure easily given the constraints of most paper-and-pencil exams. No researcher would suggest that states limit their standards only to what is measurable.

Nonetheless, the closer the alignment between state standards and tests, the more likely it is that teachers will focus on the desired content and students will have a fair chance to show what they know.

### Research Points

**Editor:** Lauren B. Resnick

**Managing Editor:** Chris Zurawsky

**Issue Researcher:** Andrew C. Porter

**Issue Reviewers:** Joan L. Herman, Suzanne Lane, Robert L. Linn

**Issue Writer:** Lynn Olson

**AERA Executive Director:** Felice J. Levine

**American Educational Research Association**

1230 17th Street, NW  
Washington, DC 20036

phone (202) 223-9485  
fax (202) 775-1824

ResearchPoints@aera.net  
www.aera.net

