Mismatch

When State Standards and Tests Don't Mesh, Schools Are Left Grinding Their Gears

By Heidi Glidden and Amy M. Hightower

Imagine this: Sylvia and Steve are seventh-graders in different states. They're both eager, hard-working students, and do reasonably well in school. Come springtime, they join most students across the country in taking various state assessments in (at least) reading and mathematics. You know these tests: they're the ones that teachers give to students on behalf of their state to monitor how students are doing in school. They are also used for federal accountability purposes to determine if schools and school districts are doing a good job educating students.

Sylvia and Steve have had different experiences with these assessments. For Sylvia, they're just par for the course. Sure, she'd rather be playing softball, but taking a test of the things she's been taught that year in school has become routine. No huge surprises, no big deal.

But bluntly put, Steve is dreading assessment season this year, based on the state test he had to take last year in math. Last year, he'd worked hard to learn the material he was taught. He always submitted the homework his teacher assigned and listened hard as his teacher explained the concepts of mean, median, and mode. From fractions and ratios to probability and circumference, Steve felt like he was mastering some tough sixth-grade math concepts. His teacher thought so too, giving him *A*s and *B*s all year. When springtime testing came around, he'd been ready to strut his stuff. But when he sharpened his #2 pencil and sat down to take the state test, darned if they didn't ask him about the Pythagorean Theorem and three-dimensional objects! These were things he hadn't studied and his teacher hadn't taught. Wait, wasn't his brother, an eighth-grader, studying some of this stuff? How was he supposed to know the answers now? Had someone given him the wrong test by mistake? No mistake: He just didn't have the knowledge he needed to answer the questions. So he did what anyone in this situation would do—he flipped through the exam and guessed. And he fidgeted. And he watched the clock, waiting for the uncomfortable moment to pass. He remembers the moment like it was yesterday.

What went wrong? Why did both Sylvia and Steve feel ready for the test, but only one of them was actually prepared? Here's a dirty little secret that educators know all too well: State tests and state content standards don't always match up. It's far too often assumed that what's expected, what's taught, and what's tested are cut from the same cloth. That's the way it should be. It's what advocates of standards-based education assumed. It's certainly rational, and it's something that's never even questioned by the general public once the test results come in—the results that judge students, schools, and sometimes teachers. But as it turns out, this assumption is too often untrue and a lot of things are at play behind the scenes.

As it happens, Steve's state isn't particularly clear about what it expects of students in each grade and in each subject. This puts his teachers in a guessing game about what to teach. It also has test developers guessing about what content to sample from as they design their assessments. Maybe they guess the same, and maybe they don't. But why leave it to chance?

Sylvia's state, in contrast, is more explicit about the grade-by-grade standards students are to meet. Her state doesn't direct teachers in *how* to teach or at what precise moment to introduce a particular concept, but it does set

Heidi Glidden, assistant director, and Amy M. Hightower, associate director, are assessment and accountability specialists for the AFT teachers division. This article is based on a research brief they published in July 2006.

ILLUSTRATIONS BY SERGE BLOCH

specific, helpful year-end goals for every grade and every subject. These standards are explicit enough for teachers like Sylvia's to build their curriculum around and for testing companies to know what content to draw upon for their tests.

While Steve and Sylvia are fictitious, the problem we've identified is real. Based on our research, just 11 states are like Sylvia's, with all of their reading and math tests clearly aligned to strong standards. The rest, to a greater or lesser extent, are like Steve's. In fact, nine states do not have any of their reading or math tests aligned to strong standards. The consequences are far-reaching since the results of these tests are used to make consequential, high-stakes judgments.

* * *

No Child Left Behind (NCLB) has led to the vast expansion of states' testing programs and heightened the stakes associated with testing results. Specifically in reading and math,* NCLB requires states to have grade-level standards in grades 3 to 8 and once in high school, and to annually test students in grades 3 to 8 and at least once in high school using assessments that are criterion-referenced/standards-based and aligned with the state's content area standards. The results of these assessments are used to determine if schools and districts are making adequate yearly progress. If not, NCLB imposes a series of escalating sanctions. (To learn more about NCLB, see **www.aft.org/topics/nclb/index.htm**.)

Given the fact that state standards are often deemed inadequate (see, for example, "The State of State Standards 2006" from the Thomas B. Fordham Institute; "Staying on Course" from Achieve Inc.; and "Making Standards Matter" from the American Federation of Teachers), we wondered how states are doing in developing assessment systems that meet NCLB's requirements and, therefore, can be legitimately used for accountability purposes. So we conducted a study to address two key questions. First, since (as we demonstrate in the next section) it is not possible to align a test to vague standards, are states' content standards in reading and math clear and specific? Second, for those standards that are clear and specific, is there evidence posted on states' Web sites for all to see that the state assessments are aligned with those standards?

For grades 3 to 8 and high school, we looked at all 50 states' and the District of Columbia's reading and math standards, as well as at the test specifications that the states and D.C. provide to their test developers.** Of course, we would have preferred to look directly at the actual tests, but they are confidential. Nevertheless, looking at the test specifications is the next best option; it seems highly unlikely that a test could be better aligned to the standards than the specifications upon which the test is based.

*NCLB also requires states to have science standards and, as of the 2007-2008 school year, administer science tests, but the law does not hold states accountable for their science results. Therefore, our main analysis focuses on reading and math, and we deal with science briefly in the box on page 31.

Just 11 states have all of their reading and math tests clearly aligned to strong standards. Nine states do not have any of their reading or math tests aligned to strong standards.

Our first step was to examine the strength, clarity, and specificity of the standards themselves. Content standards are at the heart of everything that goes on in a standards-based system, including testing. They define our expectations for what's important for children to learn, and serve as guideposts about what content to teach and assess. These state-developed public documents are the source that teachers, parents, and the general public consult to understand content-matter expectations. Content standards should exist for every single grade, kindergarten through high school, in every subject. Grade-by-grade content standards increase the likelihood that all students are exposed to a rigorous, sequenced curriculum that is consistent across schools and school districts. Grade-specific standards also make it possible to align not only assessments, but also curriculum, textbooks, professional development, and instruction. States that organize their standards grade-by-grade are best able to specify what students should learn and when they should learn it.

**For brevity's sake, throughout this document when we refer to the states collectively, we are actually referring to the 50 states and the District of Columbia.

We examined each state's content-standards documents to determine whether there was enough information about what students should learn to provide the basis for teachers to develop a common core curriculum and for the test developer to create aligned assessments. There is no perfect formula for this; we made a series of judgments based on a set of criteria. **To be judged "strong," a state's content standards had to:**

- Be detailed, explicit, and firmly rooted in the content of the subject area so as to lead to a common core curriculum;
- Contain particular content:
  - Reading standards must cover reading basics (e.g., word attack skills, vocabulary) and reading comprehension (e.g., exposure to a variety of literary genres);
  - Math standards must cover number sense and operations, measurement, geometry, data analysis and probability, and algebra and functions;
- Provide attention to both content and skills; and,
- Be articulated without excessive repetition in both math and reading in grades 3, 4, 5, 6, 7, 8, and once in high school.

For any standard we found to be strong, we then examined the extent to which the state's test specifications were aligned with the standard. In our alignment review, each state received a yes/no judgment for each of the NCLB-related tests it administered. **To meet our criteria for alignment, a state must:**

- Have evidence of the alignment of its tests and content standards through documents such as item specifications, test specifications, test blueprints, test development reports, or assessment frameworks; and,
- Post the alignment evidence on its Web site in a transparent manner.

The need for alignment should be obvious, but the need for transparency may not be. Transparency "demystifies" how (or if) the pieces connect to function as a unified *system*. A transparent system is not necessarily an aligned system, but only with transparency can we determine if the tests and content standards are aligned. A transparent testing program provides information to parents, students, teachers, and the public about the development, purpose, and use of state tests. It also brings any problems within the testing program to light so that they can be addressed. This is why, in our review, states could not simply assert that their tests were aligned to their standards. And yet, our alignment criteria were still not as stringent as we believe they should be. A state could receive alignment credit for fairly minimal documentation. For example, if a state had grade-by-grade math standards organized by number sense, algebra, measurement, etc., we gave that state credit for evidence of alignment if it indicated the percentage of items devoted to each of these topics.

Grade-by-grade content standards increase the likelihood that all students are exposed to a rigorous, sequenced curriculum that is consistent across schools and school districts.

As our opening vignette indicates, what we found was not what the average person would assume. There were two basic problems: Standards that were too weak to guide teachers or test developers, and standards that were strong, yet mismatched with tests nonetheless. To explain the problems with the weak standards, in the following section, we provide examples of vague and repetitious standards—and examples that show why tests cannot be aligned with such weak standards. We wrap up that section with data on how widespread weak standards are. Then we turn to the mismatch between strong standards and test specifications. Once again we provide examples of the mismatch as well as data on how widespread this problem is.

Vague Standards Inevitably Lead to Mismatch

The quality of content standards matters greatly to teaching, learning, and testing, so it directly affects the fairness and validity of tests and the accountability systems they support. Despite this obvious and indisputable fact, we found that across the country, many states have failed to write clear and specific standards for every subject and grade. As you read the examples† of vague state standards in the table below, consider them from both the teachers' and the test developers' perspectives. None of these standards gives enough information to teachers about what to teach or to test developers about what to test.

Subject	Grade(s)	Examples of Vague Content Standards
Reading	4	Demonstrate the understanding that the purposes of experiencing literary works include personal satisfaction and development of lifelong literature appreciation.
	8	View a variety of visually presented materials for understanding of a specific topic.
Math	4	Students will describe, extend, and create a wide variety of patterns using a wide variety of materials (transfer from concrete to symbols).
	9-12	Model and analyze real-world situations by using patterns and functions.

In contrast, take a look at the following standards; they are clear and specific enough to eliminate the guesswork.

Subject	Grade	Examples of Strong Content Standards
Reading	4	Distinguish between cause and effect and between fact and opinion in informational text. Example: In reading an article about how snowshoe rabbits change color, distinguish facts (such as snowshoe rabbits change color from brown to white in the winter) from opinions (such as snowshoe rabbits are very pretty animals because they can change colors).
Math	4	Subtract units of length that may require renaming of feet to inches or meters to centimeters. Example: The shelf was 2 feet long. Jane shortened it by 8 inches. How long is the shelf now?

† When providing examples, we chose not to name the states in the main article because it would unfairly place emphasis on them instead of on the broader problem. The examples are drawn from the following states: 1) vague standards—Arkansas, Connecticut, and Montana; 2) strong standards—Indiana; 3) repetitious standards—Connecticut and Texas; 4) mismatched standards and test specifications—Florida, Kansas, Minnesota, Montana, and Pennsylvania.

These latter examples are particularly strong—most states do not have standards this clear and specific. Instead, most states occupy a middle ground between these and the terribly vague standards shown previously. But even with middling standards, it's very hard for a teacher to know what to teach and a test developer to know what to test. Teachers may feel like they just have to make do—but test developers often do not. In states with weak standards, additional information is often given to testing companies that further clarifies or elaborates on the standard to be tested. In essence, these states are creating an additional layer or set of "shadow" standards, which are often more specific and detailed than the official standards from which they presumably came. However, it is the test developer who receives these "shadow" standards, not teachers.

Surprised? So were we. Let's look at an example to make this a little easier to understand. Here is a 4th-grade math standard and the corresponding test specification. Clearly, the test developer received much more specific information than teachers—information that would be just as helpful in preparing lessons as it is in preparing tests.

What 4th-grade teachers receive:
Describe, model, and classify two- and three-dimensional shapes

What the test developer receives:
Students demonstrate understanding of two- and three-dimensional geometric shapes and the relationships among them. In the grade 4 test, understanding is demonstrated with the following indicators as well as by solving problems, reasoning, communicating, representing, and making connections based on indicators—

- Using properties to describe, identify, and sort 2- and 3-dimensional figures [Vocabulary in addition to that for grade 3: polygon; kite; pentagon; hexagon; octagon; line; line segment; parallel, perpendicular, and intersecting lines]
- Recognizing two- and three-dimensional figures irrespective of their orientation
- Recognizing the results of subdividing and combining shapes, e.g., tangrams
- Recognizing congruent figures (having the same size and shape) including shapes that have been rotated

Clearly, it is possible for a teacher to believe she has covered a vague standard, and for a test developer to come up with an angle that she hasn't considered. In the example above, a teacher may do several lessons on describing, modeling, and classifying two- and three-dimensional shapes—but she may not think to teach students to recognize them "irrespective of their orientation," as the test specifications state. The only way to avoid such problems

Some states are creating "shadow" standards, which are often more specific and detailed than the official standards. However, it is the test developer who receives these "shadow" standards, not teachers.

is for the teachers and the test developers to receive the same clear, detailed standards.

Repetition Makes Standards Vague

Even when states manage to write standards that sound reasonably specific, they sometimes poison the effort by repeating the standard over four or more grades. This problem is especially evident in states' reading standards. For example, one state's reading standards expect eighth-graders to, among other things, "develop a critical stance and cite evidence to support the stance;" "use phonetic, structural, syntactical, and contextual clues to read and understand words;" and "describe how the experiences of a reader influence the interpretation of a text." That may sound reasonable—but the exact same thing is expected of 2nd-graders, 10th-graders, and students in every other grade in between.

Repetition of standards makes it hard, if not impossible, for a teacher to know what content students have mastered in previous grades or to determine the specific differences in student expectations from grade to grade. It certainly isn't enough for a teacher to build his or her lesson plans.

Let's look a little more at that state that expects 2nd- through 10th-graders to develop a critical stance. The vast majority of its reading standards are exactly the same from grade 3 to grade 10 and, shockingly, *more than 40 percent of the 10th-grade standards come from grade 2 standards*:

- 71 percent of the 4th-grade standards are repeated (56 percent come from grade 2)
- 87 percent of the 6th-grade standards are repeated (44 percent come from grade 2)
- 92 percent of the 8th-grade standards are repeated (42 percent come from grade 2)
- 81 percent of the 10th-grade standards are repeated (42 percent come from grade 2)

One can easily imagine how 2nd- and 9th-grade teachers, for example, would develop different lesson plans based on these repetitive standards. But what would prevent 2nd- and 3rd-grade teachers from teaching almost identical lessons? And what happens to the unlucky student who is assigned in 4th, 5th, and 6th grades to use *Charlotte's Web* to "describe how the experiences of a reader influence the interpretation of a text." Or the unlucky student who is never assigned *Charlotte's Web* for any reason? A central purpose of state standards is to avoid such repetition and such gaps—but repetitive standards that do not specify what should be taught at each grade can't serve that purpose and, as a result, they can't be used to develop standards-based tests either.

Unfortunately, the example we've been using is a pretty typical one. Here's an example of reading standards from another state that are even more repetitious from grade to grade:

- 75 percent of the 3rd-grade standards are repeated from K-2
- 98 percent of the 5th-grade standards are repeated from grade 4
- 94 percent of the 7th-grade standards are repeated from grade 4

Repetitious standards are neither clear nor specific enough to guarantee that what's taught in each and every grade and subject is also what's tested. The result? Guesswork on the part of teachers and testing companies. Or, as we saw with the vague standards, sometimes the teachers are left to guess, but the test developers get the extra information they need.

In this example, 3rd- and 4th-grade teachers work from the exact same reading standard, with no indication of what is appropriate for a 3rd-grader versus a 4th-grader. The test developer, however, receives the standard *plus* specific indicators of what is appropriate for a 3rd-grader and what is appropriate for a 4th-grader:

What 3rd- and 4th-grade teachers receive:
Determines meaning of words through knowledge of word structure (e.g., compound nouns, contractions, root words, prefixes, suffixes)

What the test developer receives:
Determines meaning of words through knowledge of word structure (e.g., compound nouns, contractions, root words, prefixes, suffixes)

Grade 3 test
Assessment Indicators
Prefixes: *mis-*, *pre-*, *pro-*, *re-*, *un-*
Suffixes: *-ed*, *-er*, *-est*, *-ing*, *-ly*, *-y*
Only test prefixes and suffixes listed above

Grade 4 test
Assessment Indicators
Prefixes: *anti-*, *dis-*, *ex-*, *non-*, *under-*
Suffixes: *-en*, *-ful*, *-less*, *-ment*, *-ness*
Only test prefixes and suffixes listed above

For example, while 3rd- and 4th-grade teachers work from the exact same standard, the test developer receives specific indicators of what is appropriate for a 3rd-grader and what is appropriate for a 4th-grader.

Unlike teachers' information about the reading standard for grades 3 and 4, the test developers receive indicators that are unique to each grade. The indicators add information that would be useful to teachers, but teachers don't receive them—nor do they necessarily know that such an elaboration even exists. An excellent 3rd-grade teacher could, in good conscience and with good reason, deliver highly effective instruction on the prefixes *anti-*, *dis-*, and *non-*, but because she guessed wrong as to what would be on the 3rd-grade test versus the 4th-grade test, her test results would indicate that her students did not know anything about prefixes. Of course, the 4th-grade teacher is in an equally difficult position—how is she to know which prefixes the students have already learned and which will be tested?

Vague and repetitious standards are clearly a big problem, but just how widespread are they? It depends on the subject. States tend to have fairly good math standards, but weak reading standards. Here is what we found:

■ **A majority of states have grade-by-grade reading and math standards in every grade that NCLB requires them to assess.** Six states still have not developed grade-by-grade standards in reading and math despite being required to do so by the guidance written for NCLB: Colorado, Illinois, Montana, Nebraska, Pennsylvania, and Wisconsin. At the high school level, 20 states clustered their reading standards and 22 clustered their math standards.

■ **But, grade-by-grade standards do not guarantee clear, specific standards: Only a little more than one-third of states have strong reading and math standards in every grade that NCLB requires them to assess.** Just 18 states and the District of Columbia met our criteria for having strong standards in reading and math in all grades that NCLB requires states to assess: California, Georgia, Indiana, Louisiana, Massachusetts, Michigan, Nevada, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, South Dakota, Tennessee, Virginia, Washington, and West Virginia.

■ **Across states and subjects, of all the 714 content standards reviewed, 70 percent met our criteria for being strong.** States had strong standards in mathematics: Eighty-seven percent of the math standards we reviewed met our criteria. In contrast, only about half of the states' reading content standards met our criteria (53 percent).

■ **On average, the most vague and repetitious content standards are in reading.** Only 20 states had strong reading standards in grades 3 to 8 and high school; 12 states had weak reading standards in all of these grades. Twenty-one percent of all reading standards reviewed were significantly repetitious across the grades (meaning word-by-word repetition across the grades at least 50 percent of the time). Fifteen states had reading standards that repeated the same reading standards in three or more grades.

(Continued on page 32)

Science Standards and Tests Suffer from Mismatch, Too

No Child Left Behind (NCLB) is somewhat more lenient with science than it is with reading and math. Science standards need not be grade by grade; academic expectations at each of the three grade-level ranges (such as grades 3 to 5, 6 to 9, and 10 to 12) are sufficient. Likewise, starting in the 2007-2008 school year, science must be assessed annually, but just once during elementary, middle/junior high, and high school—and the results are not incorporated into federally required accountability determinations.

Nonetheless, we still wanted to examine states' science standards and the extent to which their standards and test specifications are aligned. Unfortunately, as with reading and math, we found serious problems.

As we explained in the main article, grade-by-grade standards are essential for guiding instruction. And yet, 13 states cluster their science standards at the elementary level, 13 states at the middle-school level, and 21 states at the high-school level. While permitted under NCLB, clustering results in vague standards such as these:

- Grades 5 to 8—Describe the historical and cultural conditions at the time of an invention or discovery, and analyze the societal impacts of that invention;

- Grades 9 to 12—Analyze the impacts of various scientific and technological developments.

Besides getting frustrated, what is a teacher or a test developer to do with such a directive? The teacher can guess what will be tested, and the test developer can guess what will be taught. Or, they can demand more specifics from the state. For the test developers at least, such demands appear to be working.

Take a look at the following example of one 7th-grade science standard and the corresponding test specification—it reveals something we reported on in the main article with reading and math. The test designer gets the same standard that is given to teachers, as well as very specific examples that help clarify the focus of the standard.

What 7th-grade teachers receive:
The student will cite examples of individuals throughout history who made discoveries and contributions in science and technology.

What the test developer receives:
The student will cite examples of individuals throughout history who made discoveries and contributions in science and technology.

- Examples of individuals (and some of their discoveries or contributions) are limited to: Rachel Carson–*Silent Spring*; George Washington Carver–agricultural products, technology; Nicolas Copernicus–Copernican revolution; Charles Darwin–classification, ecology, and natural selection; Galileo Galilei–gravity and telescopes; Jane Goodall–primate research; James Hutton–geology; Anton van Leeuwenhoek and Robert Hooke–microscopy; Johann Gregor Mendel–genetics; Isaac Newton–gravity, mechanics, light, and telescopes; Louis Pasteur–pasteurization; and Alfred Wegener–plate tectonics.

As a teacher, wouldn't you feel like you covered the standard if you taught your students about Thomas Edison's light bulb, Eli Whitney's cotton gin, and Lord Kelvin's Kelvin scale? You might feel good, but you would not have prepared your students for a test that focused on Rachel Carson, George Washington Carver, and Johann Gregor Mendel. Teachers (and their students) would benefit significantly from the additional information provided to the test developers, but that information is not included as a part of the standards. Teachers wouldn't even know to look for this elaboration.

—H.G. and A.H.

In some states, the clarity and specificity of the standards are not the problem. The grade level and subject content to be taught are specific enough, but the tests simply cover other things.

Even with Strong Standards, Mismatch Can Happen

In some states, the clarity and specificity of the standards are not the problem; instead, it is the lack of follow-through. The grade level and subject content to be taught are specific enough, but the tests simply cover other things. For example, in one state, the 3rd-grade test pulls content from both the 3rd- and 4th-grade standards:

What 3rd-grade teachers receive:
Third-grade student uses a variety of strategies to determine meaning and increase vocabulary (for example, prefixes, suffixes, root words, less common vowel patterns, homophones, compound words, contractions)

What 4th-grade teachers receive:
Fourth-grade student uses a variety of strategies to determine meaning and increase vocabulary (for example, multiple meaning words, antonyms, synonyms, word relationships, root words, homonyms)

What the 3rd-grade test developer receives:
Third-grade test content limit—Vocabulary words for prefixes (e.g., *re-, un-, pre-, dis-, mis-, in-, non-*), suffixes (e.g., *-er, -est, -ful, -less, -able, -ly, -or, -ness*), root words, multiple meanings, antonyms, synonyms, homophones, compound words, and contractions should be on grade level

A 3rd-grade teacher in this state is unlikely to have her students prepared for questions relating to words with multiple meanings, antonyms, or synonyms because, according to the state's content standards, these concepts are not to be addressed until grade 4. As the example above demonstrates, the specific content standards that teachers receive from their state don't always match up with what the state gives test developers to create the tests.

Here's another example (taken from a different state) that reveals a similar problem. In this case, there are 8th-grade math standards and test specifications that *almost* match up. Both the standards and test specifications are about measurement, but they diverge in two important ways. First, although the standards say nothing explicitly about converting measurements, the test specification expects students to make several different types of conversions. Second, one of those conversions—moving from Fahrenheit to Celsius—involves content not even included in the 8th-grade standards.

What 8th-grade teachers receive:
Under the header "Measurement and Estimation" are the following seven standards:

- Develop formulas and procedures for determining measurements (e.g., area, volume, distance)
- Solve rate problems (e.g., rate × time = distance, principle × interest rate = interest)
- Measure angles in degrees and determine relations of angles
- Estimate, use and describe measures of distance, rate, perimeter, area, volume, weight, mass, and angles
- Describe how a change in linear dimension of an object affects its perimeter, area, and volume
- Use scale measurements to interpret maps or drawings
- Create and use scale models

What the 8th-grade test developer receives:
Assessment Anchor: Demonstrate an understanding of measurable attributes of objects and figures, and the units, systems, and processes of measurement.

Convert measurements: Eligible Content

- Convert among all metric measurements (milli, centi, deci, deka, kilo using meter, liter, and gram)
- Convert customary measurements to 2 units above or below the given unit (e.g., inches to yards, pints to gallons)
- Convert time to 2 units above or below a given unit (e.g., seconds to hours)
- Convert from Fahrenheit to Celsius or Celsius to Fahrenheit

The 8th-grade standards have content that would require students to have, as the assessment anchor requires, "an understanding of measurable attributes of objects and figures, and the units, systems, and processes of measurement." However, since teachers do not receive the specifics that the test developer receives, the 8th-grade teachers do not know to devote extra time to conversions, and the 8th-grade teachers—and their students—end up with the blame when the students perform poorly on the test.

Because of NCLB's testing requirements, states have rushed to establish tests that comply with the law. However, there appears to be very little urgency to align those tests with the content standards or be transparent about which standards are assessed. Here is what we found:

■ **Eleven states met our criteria for having both strong reading and math standards and documenting in a transparent manner that their tests align to them in all NCLB-required grades.** They are: California, Indiana, Louisiana, Nevada, New Mexico, New York, Ohio, Tennessee, Virginia, Washington, and West Virginia. Eleven states is not a lot, but keep in mind that states could fall short for several reasons—having some content standards that are weak, not aligning their strong standards to their tests, and/or not providing evidence of alignment online. Of those who fell short (39 states plus the District of Columbia), 17 did so because at least some of their testing documents were not online, 32 did so because at least some of their standards were weak, and 18 did so because their standards and tests were not aligned.

Where and Why Does Mismatch Exist?

Only 11 states met our criteria for having tests transparently aligned to strong standards: Calif., Ind., La., Nev., N.M., N.Y., Ohio, Tenn., Va., Wash., and W.Va. This table shows why the others fell short.

State	Some test specifications not online	Some mismatch between standards and test specifications	Percentage of strong reading and math standards	Percentage of tests transparently aligned to strong reading and math standards
Alabama		✓	79	64
Alaska			79	79
Arizona			71	71
Arkansas	✓		79	0
Colorado		✓	14	14
Connecticut		✓	50	0
Delaware	✓		50	0
D.C.	✓		100	0
Florida			64	64
Georgia		✓	100	57
Hawaii	✓		50	0
Idaho		✓	57	50
Illinois		✓	0	0
Iowa	✓		0	0
Kansas			50	50
Kentucky			57	57
Maine	✓		50	7
Maryland	✓		57	57
Massachusetts	✓		100	43
Michigan	✓		100	43
Minnesota			50	50
Mississippi		✓	86	79
Missouri	✓	✓	50	0
Montana	✓	✓	0	0
Nebraska	✓	✓	29	29
New Hampshire			50	50
New Jersey	✓	✓	100	43
North Carolina	✓	✓	100	43
North Dakota	✓		100	0
Oklahoma			86	86
Oregon			71	71
Pennsylvania		✓	57	57
Rhode Island			50	50
South Carolina	✓	✓	64	14
South Dakota		✓	100	50
Texas			57	57
Utah		✓	71	50
Vermont	✓		57	57
Wisconsin		✓	21	0
Wyoming		✓	71	0

■ **An additional three states had at least 75 percent of their tests aligned to strong content standards.** With a few adjustments in particular grades or in just one subject, these additional three states would fully meet our criteria for alignment to strong content standards: Mississippi (meeting 86 percent of our criteria), Oklahoma (meeting 86 percent), and Alaska (meeting 78 percent).

■ **Twice as many states met our criteria for having strong and transparently aligned standards and tests in math than they did in reading.** Twenty-six states have aligned math tests across all grades tested. But, just 13 states have aligned reading tests across all grades tested.

(Continued on page 50)

Mismatch

(*Continued from page 33)*

Overall, our results lead us to conclude that states are doing a better job in developing content standards than in using them to drive assessment. Simply put, in too many cases, tests that are not aligned to strong standards are driving many accountability systems. In order to comply with NCLB, states have been under enormous pressure to quickly develop new assessment systems. We hope this research provides some ideas on how they could improve those systems in the near future. For example, state departments of education need to post their content standards on their Web sites, along with information about how their state tests are aligned to these standards—they also need to keep this information current. When test developers or state officials clarify standards in order to write test items that align to them, the clarifications should be made public and should make their way back to the original standards document in the form of clearly marked revisions. This way, educators will be able to skip the guessing game and teach the content that the state believes is most important.

Detailed information about content standards and what will be tested should be readily available to anyone (teachers, students, parents, the general public) at any point, and should not have to be ferreted out. Educators, in particular, need to know that what will be tested draws from the content standards to which they are teaching. Where there's a mismatch, or a fuzzy match, or only an assumed match between the content that's expected and the content that's assessed—and when the results are used to judge students, schools, and teachers—it's no wonder that folks in schools toss up their hands in frustration. ❑