



Smart Testing: Let's Get It Right

How assessment-savvy have states become since NCLB?

INTRODUCTION

No one knows the value of testing better than teachers. It's the tool they use all the time to gauge how students are doing. Good teachers constantly assess students' understanding and use what they learn to tweak instruction. Testing is as integral to a teacher's stock and trade as a tape measure is to a carpenter. But today in many places teachers find themselves in an environment where standardized testing is so dominant it's distorting the relationship between teaching and learning: raising test scores is a more common topic than supporting learners; generating right answers has grown more central than the thinking behind them; and recalling data tidbits is more highly valued than developing comprehensive knowledge.

Today, testing has taken center stage in education reform, in large part as a by-product of the federal government's No Child Left Behind Act (NCLB) and its emphasis on school-based accountability. The tests that are

causing the most rumblings are states' "high stakes" exams required by NCLB. You know the ones: They're given once a year; they spark a flurry of test preparation activities to make sure students score well on them; and several months later, their results are what everyone talks about when newspaper headlines label a school as either "good" or "failing." But to what end? Teachers sometimes say these state tests don't reflect the content they are expected to teach in their classrooms, and the test results don't make their way back to teachers in a timely or user-friendly manner to be of help in adjusting instruction. And, educators often haven't had the professional development necessary to interpret the bits and pieces of assessment data they do receive to improve learning.

Let us be clear: When done appropriately, annual state-administered tests provide useful feedback about student learning and can guide the system to ensure that schools, teachers, and staff get the supports they need to help all students meet academic expectations. For

decades, the American Federation of Teachers (AFT) has embraced state-level assessment as a crucial component of a standards-based education system. But we've always insisted on a fundamental requirement: *These exams must test what the state expects teachers to teach and students to learn as part of their state standards.* Where there's a mismatch between the content that's expected, the content that's taught, 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. Without strong, clear state content standards and tests aligned to them, state-level testing is compromised and results are suspect. Unfortunately, this crucial alignment is too often assumed to be in place by politicians and pundits eager for bottom-line results.

The AFT continues to call on states and districts to administer tests that are fair, transparent, and aligned to clear, specific, and rigorous state content standards. We also feel strongly that assessment programs should be efficient and not spawn redundant, duplicative testing within the system. What we want are assessment systems that are useful to educators and that provide them with information about their students so that they can tailor instruction to most effectively meet their students' needs.

The AFT calls this “smart testing.” Smart testing is concerned with what is tested and why, whether the testing instruments are up to the task, and how test results are used. Smart testing provides information in a timely fashion, and uses results to make accurate generalizations drawn from the data. Smart testing assesses the effectiveness of the curriculum, but doesn't drive it; it informs professional development, but doesn't prescribe it; it provides information to improve teaching and learning, but in and of itself cannot cause improvement. Smart testing starts with strong, grade-specific state content standards, and includes a number of interrelated pieces (see Figure 1):

- Well-developed grade-by-grade curricula that go beyond the limited subjects required by NCLB (i.e., reading, math, and science);
- Assessments aligned to content standards;
- An efficient, valid, and reliable testing system that doesn't duplicate testing across education system levels;
- Appropriate inclusion of English language learners (ELLs) and students with disabilities in testing programs;
- Timely provision of user-friendly testing results;
- Supportive professional development, including coverage of what the content standards are and how they relate to state curricula and assessments, how to teach to the content standards, and how to use testing data to inform instruction;
- Accountability for results; and,
- Transparency of the system.

This report is concerned with three of these smart testing components: (1) strong state content standards, (2) alignment between standards and tests, and (3) transparency of the system.

While much has been written about standards and alignment, less has been said about transparency and what this means. 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. Knowledge about how the tests are developed,

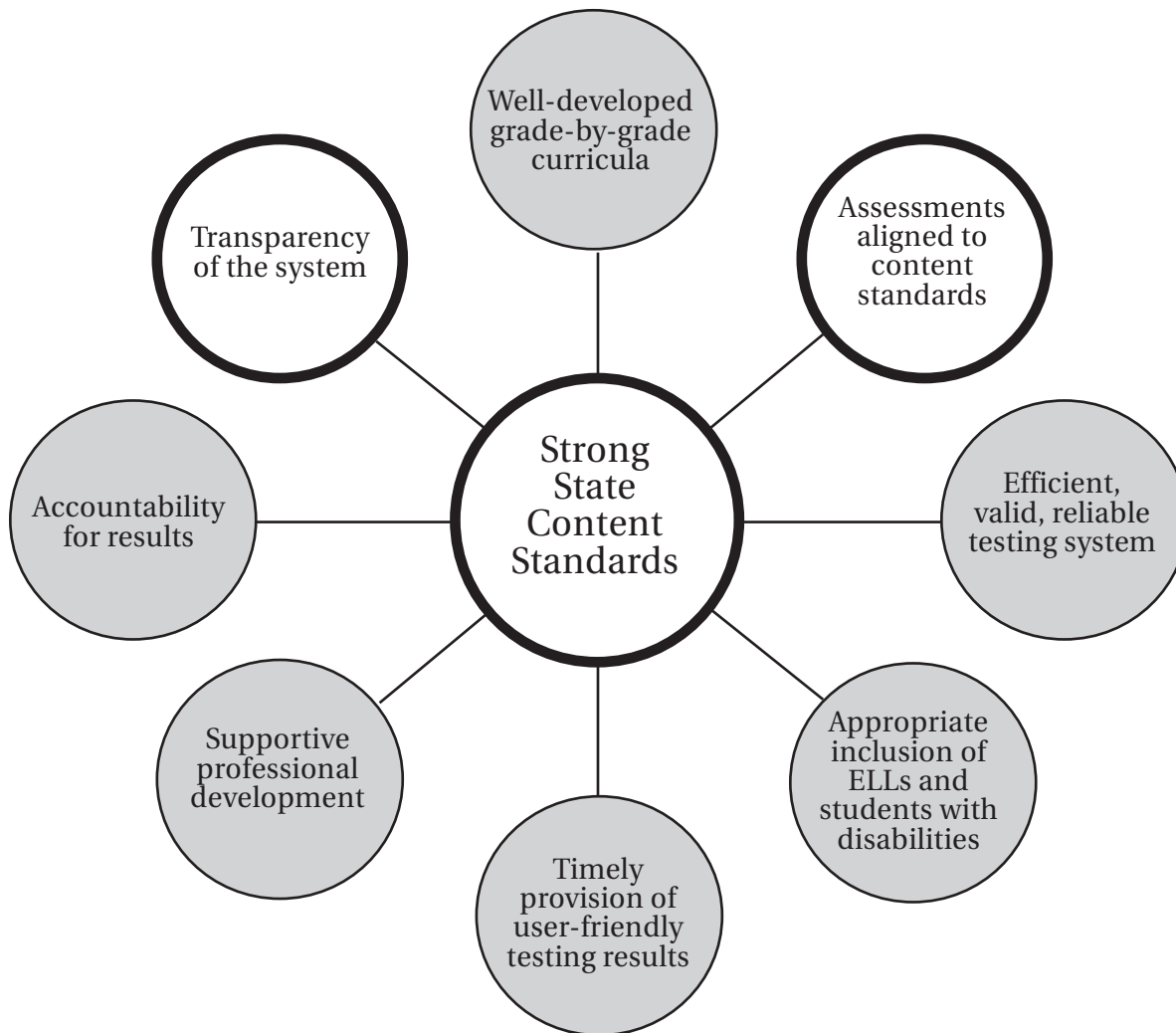
who is involved in the process, how the standards are measured, how the cut-scores are set, and how student achievement data are reported are just some of the crucial details a transparent system provides. Moreover, a transparent system brings any problems within the testing program to light so that they can be addressed. By providing information to the public and helping to ensure responsiveness to problems and concerns, transparency helps build professional and public acceptance and trust in the testing program and accountability measures.

The AFT has a long history of analyzing these smart testing components as a part of our advocacy for standards-based reform. From 1995 to 2001, the AFT published an annual report, *Making Standards Matter (MSM)*,

which judged states' efforts to institute a standards-based education system, including the evolution of their testing programs and their alignment to standards. Last published just as NCLB came into law, *MSM* found that without mid-course corrections the standards movement was headed for trouble. We called for stronger curriculum to support state content standards, higher-quality tests aligned to strong and rigorous state standards, and additional funds to assist students who fail to meet the standards.

This report revisits the status of large-scale state assessment in light of NCLB's emphasis on test-based accountability and the repeated concerns AFT members have expressed about testing's unintended consequences.

Figure 1: Smart Testing



Our members frequently tell us testing has taken on a life of its own, too often separate and apart from the standards and curriculum it was intended to support. Mounting evidence from across the country suggests that states and the federal government have rushed to develop assessments and accountability and overlooked (or, in some cases, ignored) both the process of assessing the quality and rigor of the standards, or of aligning curriculum and professional development to them, or of getting extra support to the students, educators, and schools that most need it.

NCLB led to the vast expansion of states' testing programs and heightened the stakes associated with testing results. Have states become smarter about assessment or overwhelmed by the mounting complexity of the task? Have the problems we identified in 2001 been resolved, or are these same flaws now perpetuated, even amplified? What new challenges have developed as the importance of state tests has grown? It's time for another look at statewide testing programs to see what progress has been made or what ground has been lost in the past five years. Smart testing: It's high time to get it right.

CRITERIA AND METHODS

This report updates AFT's 2001 *MSM* findings in two key areas: the alignment of state tests to state content standards and the strength of these content standards. We looked at all 50 states and the District of Columbia for this review.¹ The tests that were our focus are the large-scale, summative assessments that states are required by federal law to administer in three specific content areas: reading, mathematics, and science. To receive federal education funds under NCLB, states must: (1) have developed grade-level expectations in reading and math, and at each of the three grade-level ranges (typically grades 3-5, 6-9, and 10-12) in science; (2) annually assess students in grades 3-8 and at least once in high

school in reading and math; and (3) starting no later than 2007-08, annually assess students in science at least once during elementary, middle/junior high, and high school (about three-quarters of states test science now). NCLB requires that assessments be criterion-referenced/standards-based and aligned with the state's content area standards. Thus, we focused our review on an analysis of the content standards related to these three NCLB-required testing areas.

Since the AFT's push for standards-based education and the initial publication of *MSM* in 1995, all states have developed content standards in core subject areas and have implemented testing in reading and math for grades 3-8 and once in high school. This report addresses two questions:

- Are the content standards clear and specific for each subject and grade tested?
- Are the state assessments in reading, math, and science (in states already assessing science achievement) aligned with strong standards?

How We Judged Content Standards

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 grades, schools, and school districts. Grade-specific standards also facilitate greater alignment of

¹ D.C. is counted as a state in our breakdowns and analysis.

Table 1: Examples of Strong and Weak Standards

| | Strong Standards | Weak Standards |
|--------------------|--|--|
| Reading | Students should apply knowledge of word origins, derivations, synonyms, antonyms, and idioms to determine the meaning of words and phrases. <i>(Grade 4)</i> | Students should be able to construct meaning through experiences with literature, cultural events, and philosophical discussion. <i>(No grade level indicated)</i> |
| Mathematics | The student will differentiate between area and perimeter and identify whether the application of the concept of perimeter or area is appropriate for a given situation. <i>(Grade 5)</i> | Students should become mathematical problem solvers. To develop these abilities, students need the experience of working with diverse problem-solving situations. <i>(No grade level indicated)</i> |
| Science | Students should be able to describe the basic process of photosynthesis and respiration and their importance to life. <i>(Grade 5)</i> | Students should be able to use basic science concepts to help understand various kinds of scientific information. <i>(Upper elementary)</i> |

standards-based curriculum, assessments, textbooks, 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.

The quality of content standards varies enormously from state to state, subject to subject, and grade to grade. Standards can be full of empty rhetoric, unclear, and devoid of content. They can be so vast and scattered that no teacher could prepare a student to meet them in the course of a year. If they are too vague, teachers and test developers can't hope to focus on the same materials. And, if they're too narrow, they constrict the curriculum. When standards don't make priorities clear, teachers end up in a guessing game of selecting what they will teach. Moreover, when content standards are poor, they are of little help in informing test developers on what to assess. The quality of content standards matters greatly to the interrelated functions of teaching, learning, and testing as well as to the fairness and validity of tests and the accountability systems they support.

The Criteria

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 a common core curriculum and assessments. There is no perfect formula for this; we made a series of judgment calls based on a set of criteria. To be judged "strong," a state's content standards must:

- Be detailed, explicit, and firmly rooted in the content of the subject area 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.

- Science standards must cover earth, physical, and life sciences;
- Provide attention to both content and skills; and,
- Be articulated for the following individual grades and subjects²: math—grades 3, 4, 5, 6, 7, 8, and once in high school; reading—grades 3, 4, 5, 6, 7, 8, and once in high school; and science—once at each of the three grade clusters (elementary, middle, and high school), and not contain excessive repetition across grade levels.

In general, strong content standards provide clear guidance to teachers, curriculum and assessment developers, textbook publishers, and others so that one person’s interpretation of the central knowledge and skills students should learn at a particular grade will be comparable to someone else’s. Table 1 presents examples of state standards that meet and do not meet AFT’s criteria.

What We Examined

We examined only those standards documents that states posted on their Web sites as the basis of their state tests (including relevant documents that go by the names of “content standards,” “learning standards,” “indicators,” “expectations,” “curriculum frameworks,” etc.). To be judged as having strong content standards across the board, a state had to meet our criteria for strong content standards in *each* of the 17 different content standards across NCLB-tested subjects and grades.

² While truly strong content standards would always be grade-specific or, alternatively at the high school level, course-specific, for this review we made a few exceptions to conform to the guidance states follow for NCLB. We accepted standards that are clustered by grade level for science and at the high school level for all subject areas.

How We Judged Whether Tests Are Aligned to Standards

If a state wants to know if students are meeting its content standards or if a school is effective in helping students meet the standards, a state’s test must sample from the content specified in its standards. We cannot know if students or schools are meeting expectations when the tested material is not specified in the standards that teachers have been asked to teach. It is equally unfair to hold students, teachers, and schools accountable on the basis of such faulty assessments. For example, if a teacher has taught and a student has studied hard all year to learn decimals and fractions, but in March is tested on knowledge of division, it is not an accurate assessment of the student’s year-long efforts or of the school’s effectiveness.

The Criteria

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 strong content standards;
- Provide evidence of the alignment of its tests to its content standards, i.e., item specifications, test specifications (“test specs”), test blueprints, test development reports, or assessment frameworks; and,
- Post the alignment evidence on its Web site in a transparent manner.

States could not simply assert that their tests were aligned to their standards. They also had to provide alignment documentation. While we prefer documentation that goes below the strand level (e.g., math standards that are organized by number sense, algebra, measurement, etc.), we gave a state credit for evidence of alignment if it indicated the percentage of items devoted to each strand.

What We Examined

Across the 50 states and the District of Columbia, we reviewed 17 different content standards and documents related to up to 17 different tests for each state.³ In total, we reviewed 861 content standards documents (357 in reading, 357 in math, and 147 in science) and information pertaining to 833 state tests (357 in reading, 357 in math, and 119 in science).

We only reviewed materials that were posted on state department of education Web sites in May and June 2006. While this strategy limited the scope of what we could examine, our approach was a deliberate test of transparency—information about content standards and what will be tested should be readily available to anyone (teachers, 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. They should be able to go to the state's Web site and find all relevant materials posted in a user-friendly manner. Web site accessibility ensures that teachers have immediate access to any changes and, if they are assigned to teach a new subject or grade, that the testing and alignment materials are available when they need them. For states, providing these materials on the Web is far more cost-effective and efficient than copying and distributing thousands of new standards documents for teachers and repeating this process any time a change is made.

Our analysis did not consider the technical qualities of the tests, and we did not examine the tests themselves. We also did not look at states' proficiency standards or whether the

³ Despite NCLB's requirement for science content standards to be in place by the current school year, several states were still developing science standards and related tests at the time of our review: seven science content indicators were missing; and in 37 cases, states had not yet identified which science test they will use for NCLB reporting. Because the law does not require science assessment until the 2007-08 school year, we allowed states to be exempt from this requirement in our overall judgments.

content standards were grade- or otherwise appropriate. These are all important smart testing issues, but they were beyond the scope of this particular report. We also limited our review to the three content areas required by NCLB: reading, mathematics, and science.⁴

It is important to keep in mind the changes in federal law brought about by NCLB and the corresponding evolution of AFT's criteria. No Child Left Behind put more tests and grade levels into play, enormously compounding the complexity of assessment and accountability for states. Thus, NCLB has set a much higher bar for states in terms of developing grade-level expectations, aligning a greater number of assessments to them, and making the alignment publicly transparent than when we reviewed these same features in 2001. For AFT's 2006 review, we adjusted our criteria to match more closely with these changes in federal law. For example, now that states must test annually in grades 3-8 in reading and math, for AFT to give credit for strong standards, states *must* have grade-specific content standards in reading and math to guide teaching, learning, and testing. In prior years, we preferred a grade-by-grade approach to writing content standards, but we gave states credit if they had strong content standards that were bundled by each grade cluster (usually grades 3-5 and 6-8). Today, we no longer give states credit for clustered reading or math standards at the elementary and middle levels, no matter how clear and specific they may be.

FINDINGS

States with Strong Standards and Tests Aligned to Them

In order to be strong in both content standards and alignment, a state had to meet all of our criteria for strong standards and all of our

⁴ Because of NCLB's emphasis, our analysis covers only reading, math, and science, in contrast to earlier editions of *Making Standards Matter*, which also covered writing and social studies.

Table 2: Summary of States' Progress Toward Test Alignment to Strong Content Standards

Percent of State Tests Aligned to Strong Content Standards

| 100% | 75-99% | 50-74% | 26-49% | 1-25% | 0% |
|---|---|--|---|--|--|
| 1. California 2. Indiana 3. Louisiana 4. Nevada 5. New Mexico 6. New York 7. Ohio 8. Tennessee 9. Virginia 10. Washington 11. West Virginia | 1. Alaska 2. Arizona 3. Mississippi 4. Oklahoma 5. Oregon | 1. Alabama 2. Florida 3. Georgia 4. Kansas 5. Kentucky 6. Maryland 7. Massachusetts 8. Minnesota 9. New Hampshire 10. New Jersey 11. Pennsylvania 12. Rhode Island 13. South Dakota 14. Texas 15. Utah | 1. Idaho 2. Michigan 3. Nebraska 4. North Carolina 5. Vermont | 1. Colorado 2. Illinois 3. Maine 4. Missouri 5. South Carolina 6. Wisconsin | 1. Arkansas 2. Connecticut 3. Delaware 4. District of Columbia 5. Hawaii 6. Iowa 7. Montana 8. North Dakota 9. Wyoming |

criteria for aligned assessments, or a total of up to 34 different criteria.

- **Eleven states met our criteria for having both strong content standards and documenting in a transparent manner that their tests align to them in all NCLB-required grades and subjects.**

They are: California, Indiana,* Louisiana, Nevada, New Mexico,* New York, Ohio, Tennessee, Virginia, Washington, and West Virginia.
 (*State isn't yet testing in science at the high school level.)

The AFT commends these 11 states for meeting this challenge. We consider these states to be leading the pack in terms of standards, alignment, and transparency.

Of these 11 states, Tennessee stands out. In addition to having strong standards across the board and tests aligned to them, its standards documents clearly specify which standards will be tested, and its high school standards are written course by course.

But an all-or-nothing designation does not tell the whole story, nor does it give an accurate picture of state efforts at aligning tests with

content standards. States vary in how close they are to having strong content standards and aligned tests in all grades and subjects (see Table 2).

- **Almost one-third of all states (16) had at least 75 percent of their tests aligned to strong content standards.**

Eleven states (listed above) met 100 percent of our criteria. The additional five states that are close are: Mississippi (meeting 82 percent of our criteria), Oklahoma (meeting 82 percent), Alaska (meeting 79 percent), Arizona (meeting 76 percent), and Oregon (meeting 76 percent).

With a few adjustments in particular grades and subjects, these additional five states would fully meet our criteria for alignment to strong content standards. They are in the home stretch.

- **Thirty-one states are at least halfway along in terms of providing evidence that their state assessments align with strong state content standards.**

Fifteen of these states had from 50 percent to 74 percent of their tests aligned to strong content standards.

These states are on the right track but need to step up their efforts.

- **The other 20 states have much work to do.**

Five states had from one-quarter to one-half of their tests aligned with strong content standards.

Six states had only minimal evidence of alignment to strong content standards; less than one-quarter of their tests align with strong content standards.

Nine states (more than 15 percent) did not meet our criteria for tests aligned to strong standards in any grade or subject. Most of these states failed to provide any information about alignment beyond released test items. These nine states have the furthest to go in terms of standards development, test alignment, and system transparency.

States fell short either by not providing all testing documents online, by failing to match up all test documents with standards, and/or by having some content standards that are weak, particularly in reading (see Table 3).

Findings Regarding Strength of State Content Standards

States have been required by federal law to develop content standards for more than 10 years. This is nothing new for states, and the AFT believes that states should be proficient at developing content standards by now. That said, NCLB guidance went from requiring nine different content standards per state to requiring 17; states that had weak content standards before NCLB were faced with both writing stronger content standards and developing tests of them in a very short period of time. Here is where we found states to be today in terms of the strength of their content standards (see Table 4):

- **A majority of states now have grade-by-grade content standards in all three NCLB-**

related subjects, although these standards are not universally strong.

Some states still have not developed grade-by-grade standards in reading and math despite its being required by the guidance written for NCLB: Colorado, Illinois, Montana, Nebraska, Pennsylvania, and Wisconsin. At the high school level, 20 states clustered their reading standards, 22 clustered their math standards, and 21 clustered their science standards.

- **More than one-third of states had strong standards in every grade and subject.**

Eighteen states met our criteria for having strong standards in *all* assessed grades and subjects reported under NCLB: 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.

- **Of all the 861 content standards reviewed across states and NCLB-related subjects, 74 percent met our criteria for being strong.**

States had the strongest standards in science: Ninety-five percent of the science standards we reviewed met our criteria.

States also had strong standards in mathematics: Eighty-seven percent of the math standards we reviewed met our criteria. Forty-one states had strong math standards in grades 3-8 and high school; only three states had weak math standards in all of these grades.

- **On average, the weakest content standards are still reading.**

Only 20 states had strong reading standards in grades 3-8 and high school; 12 states had weak reading standards in all of these grades.

Table 3: Where 40 States Fell Short on Tests Aligned to Strong Content Standards

| | Some testing documents not online | Some test documents do not match standards | Some standards are weak |
|----------------------|-----------------------------------|--|-------------------------|
| Alabama | | ✓ | ✓ |
| Alaska | | | ✓ |
| Arizona | | | ✓ |
| Arkansas | ✓ | | ✓ |
| Colorado | | ✓ | ✓ |
| Connecticut | | ✓ | ✓ |
| Delaware | ✓ | | ✓ |
| District of Columbia | ✓ | | |
| Florida | | | ✓ |
| Georgia | | ✓ | |
| Hawaii | ✓ | | ✓ |
| Idaho | | ✓ | ✓ |
| Illinois | | ✓ | ✓ |
| Iowa | ✓ | | ✓ |
| Kansas | | | ✓ |
| Kentucky | | | ✓ |
| Maine | ✓ | | ✓ |
| Maryland | ✓ | | ✓ |
| Massachusetts | ✓ | | |
| Michigan | ✓ | | |
| Minnesota | | | ✓ |
| Mississippi | | ✓ | ✓ |
| Missouri | ✓ | ✓ | ✓ |
| Montana | ✓ | ✓ | ✓ |
| Nebraska | ✓ | ✓ | ✓ |
| New Hampshire | | | ✓ |
| New Jersey | ✓ | ✓ | |
| North Carolina | ✓ | ✓ | |
| North Dakota | ✓ | | |
| Oklahoma | | | ✓ |
| Oregon | | | ✓ |
| Pennsylvania | | ✓ | ✓ |
| Rhode Island | | | ✓ |
| South Carolina | ✓ | ✓ | ✓ |
| South Dakota | | ✓ | |
| Texas | | | ✓ |
| Utah | | ✓ | ✓ |
| Vermont | ✓ | | ✓ |
| Wisconsin | | ✓ | ✓ |
| Wyoming | | ✓ | ✓ |
| TOTALS | 17 | 18 | 32 |

Table 4: Content Standards Meet the AFT Criteria for Being Strong

| | Reading | | | | | | | Math | | | | | | | Science* | | | % of strong standards |
|----------------------|---------|---|---|---|---|---|----|------|---|---|---|---|---|----|----------|----|----|-----------------------|
| | 3 | 4 | 5 | 6 | 7 | 8 | hs | 3 | 4 | 5 | 6 | 7 | 8 | hs | e | m | hs | |
| Alabama | ✓ | ✓ | ✓ | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 82 |
| Alaska | ✓ | ✓ | | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 82 |
| Arizona | ✓ | ✓ | | ✓ | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 76 |
| Arkansas | ✓ | ✓ | ✓ | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 82 |
| California | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| Colorado | | | | | | | ✓ | | | | | | | ✓ | | | ✓ | 18 |
| Connecticut | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 59 |
| Delaware | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 59 |
| District of Columbia | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | UD | UD | UD | 82 |
| Florida | ✓ | ✓ | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 71 |
| Georgia | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| Hawaii | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 59 |
| Idaho | ✓ | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 65 |
| Illinois | | | | | | | | | | | | | | | ✓ | ✓ | ✓ | 18 |
| Indiana | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| Iowa | | | | | | | | | | | | | | | | | | 0 |
| Kansas | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 59 |
| Kentucky | | | | ✓ | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 65 |
| Louisiana | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| Maine | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 59 |
| Maryland | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 65 |
| Massachusetts | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| Michigan | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| Minnesota | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 59 |
| Mississippi | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 88 |
| Missouri | | | | ✓ | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | 59 |
| Montana | | | | | | | | | | | | | | | | ✓ | ✓ | 12 |
| Nebraska | | | | | | | ✓ | | ✓ | | | | ✓ | ✓ | | ✓ | ✓ | 35 |
| Nevada | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| New Hampshire | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | UD | UD | UD | 41 |
| New Jersey | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| New Mexico | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| New York | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| North Carolina | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| North Dakota | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| Ohio | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| Oklahoma | ✓ | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | 82 |
| Oregon | ✓ | ✓ | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 76 |
| Pennsylvania | ✓ | | ✓ | | | | ✓ | ✓ | | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ | 65 |
| Rhode Island | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 59 |
| South Carolina | ✓ | | | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | 71 |
| South Dakota | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| Tennessee | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| Texas | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 65 |
| Utah | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 76 |
| Vermont | ✓ | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 65 |
| Virginia | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| Washington | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| West Virginia | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| Wisconsin | | | | | | | ✓ | | ✓ | | | | | ✓ | ✓ | ✓ | ✓ | 35 |
| Wyoming | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 76 |

UD = Under Development

*=NCLB required states to have developed science standards at each of the three grade-level ranges by 2005-06.

Only about half of the states' reading content standards met our criteria (53 percent).

Twenty-one percent of all reading standards reviewed were significantly redundant across the grades. By significantly redundant, we mean word-by-word repetition across the grades at least 50 percent of the time. So, although states may have had reading standards that are grade-specific, a number of them simply repeat half or more of the wording of the standard from the adjacent grade(s). Fifteen states have reading standards that repeated the same reading standards in three or more grades. This is inadequate in helping teachers or test developers know what is expected of students in reading, in each grade.

Findings Regarding Test Alignment to State Content Standards

Because testing and school accountability have received the most attention under NCLB, states have focused attention on establishing grade-by-grade tests that comply with the law. However, the urgency to align those tests with the content standards or be transparent about which standards are assessed has received comparably less attention. Today, in terms of the transparency of alignment of states' tests to content standards we find the following (see Table 5):

- **Many states continue to struggle in aligning their tests with strong content standards.**

Only 11 states fully met our criteria for alignment. Just over 20 percent of the states clearly documented how their tests are aligned in *all* assessed grades and subjects reported under NCLB.

Twenty-six states have aligned *math* tests across all grades tested.

Twenty-three states have aligned *science* tests across all grades tested.

Thirteen states have aligned *reading* tests across all grades tested.

- **Nearly one-quarter (23 percent) of states' content standards also specified the standards that will be assessed by the state.**

Fourteen states specified which math standards will be tested.

Twelve states specified which reading standards will be tested.

Ten states specified which science standards will be tested.

- **Fifty-two percent of the 833 state-administered tests met our criteria for alignment to strong content standards.**

Sixty-nine percent of all science tests were based on strong science standards.

Sixty-one percent of all math tests were based on strong math standards.

Thirty-nine percent of all reading tests were based on strong reading standards.

DISCUSSION

Our dual findings that 74 percent of the content standards across states are strong, but that only 52 percent of states' tests are aligned to strong standards, 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, testing unaligned to strong standards is driving many accountability systems. *The result: Those states' systems are not yet "smart" enough to bear the weight of the accountability functions they are asked to serve.*

There is good news...

- Overwhelmingly, states have moved to grade-by-grade content standards, which

Table 5: Tests That Meet AFT's Criteria for Being Aligned to Strong Content Standards

| | Reading | | | | | | | Math | | | | | | | Science* | | | % aligned as required for NCLB '05-'06 |
|----------------------|---------|---|---|---|---|---|----|------|---|---|---|---|---|----|----------|----|----|--|
| | 3 | 4 | 5 | 6 | 7 | 8 | hs | 3 | 4 | 5 | 6 | 7 | 8 | hs | e | m | hs | |
| Alabama | ✓ | ✓ | ✓ | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | UD | UD | | 60 |
| Alaska | ✓ | ✓ | | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | UD | UD | UD | 79 |
| Arizona | ✓ | ✓ | | ✓ | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | 76 |
| Arkansas | | | | | | | | | | | | | | | UD | UD | UD | 0 |
| California | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | 100 |
| Colorado | | | | | | | ✓ | | | | | | | ✓ | | | ✓ | 18 |
| Connecticut | | | | | | | | | | | | | | | | | | 0 |
| Delaware | | | | | | | | | | | | | | | | | | 0 |
| District of Columbia | | | | | | | | | | | | | | | UD | UD | UD | 0 |
| Florida | ✓ | ✓ | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 71 |
| Georgia | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | | | | | UD | UD | ✓ | 60 |
| Hawaii | | | | | | | | | | | | | | | | | | 0 |
| Idaho | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | 47 |
| Illinois | | | | | | | | | | | | | | | | | ✓ | 6 |
| Indiana | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | UD | 100 |
| Iowa | | | | | | | | | | | | | | | | | | 0 |
| Kansas | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 59 |
| Kentucky | | | | ✓ | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 65 |
| Louisiana | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| Maine | | | | | | | | | | | | | | ✓ | | | | 6 |
| Maryland | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 65 |
| Massachusetts | | ✓ | | | | ✓ | ✓ | | ✓ | | | | ✓ | ✓ | ✓ | ✓ | ✓ | 53 |
| Michigan | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | | | 35 |
| Minnesota | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 59 |
| Mississippi | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | 82 |
| Missouri | | | | | | | | | | | | | | | ✓ | ✓ | ✓ | 18 |
| Montana | | | | | | | | | | | | | | | | | | 0 |
| Nebraska | | | | | | | ✓ | | ✓ | | | | ✓ | ✓ | | ✓ | | 29 |
| Nevada | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| New Hampshire | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | UD | UD | UD | 50 |
| New Jersey | | | ✓ | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ | ✓ | 53 |
| New Mexico | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | UD | 100 |
| New York | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| North Carolina | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | | | | | | | UD | UD | UD | 43 |
| North Dakota | | | | | | | | | | | | | | | UD | UD | UD | 0 |
| Ohio | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| Oklahoma | ✓ | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | 82 |
| Oregon | ✓ | ✓ | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 76 |
| Pennsylvania | ✓ | | ✓ | | | ✓ | ✓ | ✓ | ✓ | | | | ✓ | ✓ | ✓ | | | 53 |
| Rhode Island | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | UD | UD | UD | 50 |
| South Carolina | ✓ | | | ✓ | | | | | | | | | | | UD | UD | ✓ | 20 |
| South Dakota | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | | | | | | UD | UD | UD | 50 |
| Tennessee | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| Texas | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 65 |
| Utah | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 59 |
| Vermont | ✓ | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | | 47 |
| Virginia | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| Washington | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| West Virginia | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100 |
| Wisconsin | | | | | | | | | | | | | | | ✓ | ✓ | ✓ | 18 |
| Wyoming | | | | | | | | | | | | | | | | | | 0 |

UD = Under Development

* = Science testing under NCLB isn't required until 2007-08

if well-developed, are helpful in guiding instruction and test development.

- Overall, the content standards are more specific than the ones we reviewed in 2001. This is particularly notable given that states have 17 different content standards to write today, versus the nine required five years ago.
- States' Web sites are the source of all of the information we reviewed. States have made tremendous strides in this area. The standards and tests based on them are more transparent than five years ago.

There is bad news...

- Although states have made substantial progress in what they are posting on their Web sites, too often the information provided is out of date and/or unnecessarily scattered and buried in too many places on the site.
- Some states are administering tests but do not post any information on their Web sites about which standards are assessed.
- States continue to struggle to articulate strong reading standards, including helpful descriptions of what students need to know at each grade level.
- States also are not articulating clearly enough what high school students should be learning in reading, math, and science.
- Nearly half of the testing (48 percent) is based on weak standards and/or unaligned tests.

RECOMMENDATIONS

State departments of education are under tremendous stress from the increased testing demands and the sheer volume of data produced. While we acknowledge these pressures—which also come at a time when many states have

fewer resources and reduced staff in these departments—there are some straight-forward steps that states can take to strengthen content standards and increase the transparency and alignment of their systems:

- The overall weak quality of reading standards is alarming. Cited states need to write clearer and more specific content standards in reading at all grade levels. States need to move quickly; they should be encouraged to work together in voluntary consortia to jointly develop reading standards that better describe what students should know and be able to do in reading, grade-by-grade through high school, or look at the strong reading standards of other states.
- States that do not have grade-by-grade or course-specific standards need to develop these standards to improve the quality of their high schools. States need to write clearer and more specific content standards in all three subject areas at the high school level. Again, states should be encouraged to work with other states to articulate at a grade-by-grade or course-by-course level of specificity what high school students should know and be able to do.
- States need to move quickly toward the creation of science standards that are grade-specific, even though NCLB only requires clustered standards by grade levels. While this analysis gave states credit for clustered science standards that were clear and specific, truly strong standards in any content and across all grade levels should be written grade-by-grade so that teachers know what to teach at each grade. If states act quickly in this regard, they will be better prepared if NCLB requires grade-by-grade science testing in the future.
- To help *all* states get testing right and fulfill the goals of NCLB, Congress must go beyond the bare minimum required

by the law and appropriate additional funding for state departments of education to develop quality assessment systems. Getting testing right means not only developing strong content standards and tests based on them but also attending to the other features of smart testing identified in this report (see Figure 1).

- States must better coordinate the work of the assessment and curriculum divisions within state departments of education. All too often, we found information from one department contradicting information from another department on state Web sites. State departments should ensure that staff responsible for content and staff responsible for assessment work together to create crosswalks between the content standards documents and testing documents to show alignment; these should be placed prominently on the Web.
- 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 and 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 revisions.
- Once states get testing right, they have a responsibility to provide professional development around their assessment programs and the data they produce. Teachers and administrators should know how to use the content standards and assessment results. Teachers and paraprofessionals must have access to meaningful professional development on the best ways to use assessment results to improve instruction. The AFT-developed course for educators, “Making Data Work

for You,”⁵ is a good source for helping teachers understand, interpret, and use test results for school improvement.

The AFT also calls for an in-depth, national review of the quality of tests in use and their effects on education. Such a review should ensure that states and testing companies are strictly adhering to common, professional testing practices in the development of large-scale assessment programs and in the interpretation and use of the data generated by these assessment/accountability systems. This AFT analysis was unable to consider the statistical validity or reliability of the assessments being used, the breadth and scope of coverage from standards to assessments, or the effects of ill-designed, unaligned assessments at the classroom or student level. These are important and lingering questions that must be examined as a part of smart testing. Let’s get it right, right now.

⁵ Contact AFT’s Educational Issues Department for more information about this course.



A Union of Professionals

American Federation of Teachers, AFL-CIO
555 New Jersey Ave. N.W.
Washington, DC 20001
202/879-4400
www.aft.org