

WHERE WE STAND:
**STANDARDS-BASED
ASSESSMENT
AND ACCOUNTABILITY**



A Union of Professionals

AFT Teachers



WHERE WE STAND:
**STANDARDS-BASED
ASSESSMENT
AND ACCOUNTABILITY**

AFT Convention Resolutions

Standards-Based Assessment and Accountability / 3

Achieving the Goals of Standards-Based Reform / 7

Questions and Answers / 15

Background Information / 20

Background Reading / 27

Copyright © American Federation of Teachers, AFL-CIO (AFT 2003). Permission is hereby granted to AFT state and local affiliates to reproduce and distribute copies of this work for nonprofit educational purposes, provided that copies are distributed at or below cost, and that the author, source, and copyright notice are included on each copy. Any distribution of such materials by third parties who are outside of the AFT or its affiliates is prohibited without first receiving the express written permission of the AFT.

AFT CONVENTION RESOLUTION

Standards-Based Assessment and Accountability

Adopted July 2002

WHEREAS, because the AFT strongly supports standards-based reform, including appropriate testing, it is especially outrageous that this critical reform is threatened by uninformed implementation. The public and teachers are understandably deeply troubled that standardized tests are all too often being used inappropriately, are usurping too much instructional time, and are crowding out recognition of other important subject areas; and

WHEREAS, the public, parents and teachers remain solidly behind standards-based education reform, which testing reveals is paying off in solid and sometimes dramatic progress in raising all students' achievement; and

WHEREAS, the results of external tests, along with classroom assessments, are critical professional tools for teachers, enabling them to understand more fully their students' and their own strengths and weaknesses and thereby devise appropriate instructional and professional development strategies; and

WHEREAS, taxpayer-funded institutions must be open and accountable to the public that supports them and such openness and accountability represent key distinctions between public and private institutions; and

WHEREAS, openness and public accountability are not only two of the hallmarks of democratic institutions, but they are also necessary means for assessing and furthering progress in the achievement of democratic values and goals, such as fairness, equal opportunity and the well-being of all members of society; and

WHEREAS, in education, achievement testing and the public report-

ing of results from all schools receiving public funds are legitimate and necessary means for discharging the responsibility to be open and accountable to the public and for determining and furthering progress in the attainment of national goals such as equal educational opportunity, the elimination of achievement gaps based on family background and raising the academic performance of all students; and

WHEREAS, achievement testing and reporting are also two of the necessary means for public education to fulfill its responsibility to the parents who entrust their children to our schools, and, in turn, can help parents better participate in their children's education; and

WHEREAS, in too many instances, the public and teachers are rightly concerned that standardized tests alone, rather than standards-based teaching and learning, are increasingly driving standards-based reform. Most disturbing is that these tests are often not aligned to the state's academic standards, are based on vague or weak standards or do not measure the depth and breadth of the standards; and

WHEREAS, parents and educators alike worry that children are being over tested, taking time away from valuable instruction, and that a reliance on standardized test results encourages a "teach-to-the-test" mentality that narrows the curriculum; and

WHEREAS, very few states and local districts adhere to professional standards for test development and use, and only a handful of states use tests that are based on curricula linked to high standards, and even fewer have provided the support teachers and students need to meet the higher standards; and

WHEREAS, many states and local districts grossly misuse test results when they make high-stakes decisions affecting students, schools or school staff, based on testing and accountability systems that do not meet professional standards in testing and measurement; and

WHEREAS, it is essential that standards-based education not be hijacked by improper test use and that tests be used responsibly and in ways that support, rather than undermine, the teaching and learning process. In particular, we must separate the poor, and sometimes punitive, implementation of some testing and accountability systems from the legitimate need for professionally sound achievement testing to measure the progress of students in meeting the standards and to diagnose problems:

RESOLVED, that the AFT will:

- continue to support appropriate, high-quality testing as an essential element of standards-based reform;
- continue to oppose the abuse, misuse and overuse of standardized testing;
- call upon the federal government to create an independent, nonpartisan entity to serve as a public source of information on effective testing and accountability systems and about practices that fall short of professional standards for the development and proper use of tests;
- monitor student testing practices to ensure that tests are used responsibly and in ways that support and enhance the teaching and learning process. The AFT will track and evaluate state efforts to develop coherent standards-based systems with special emphasis on the states' testing and accountability programs;
- develop and publicize guidelines for the use of standardized assessments. Such guidelines will include information regarding the need for
 - teachers to have a voice in the development, implementation, evaluation and use of standardized tests;
 - strong standards, curriculum, extra help for struggling students the necessity for appropriate funding, and standards-based professional development to be in place before tests are used for high-stakes decisions;
 - tests to meet professional standards if they are to be used as part of high-stakes decisions;
 - administrators and state education department officials to use tests only for the purposes for which they are designed;
 - other pertinent information on student achievement, including teacher professional judgment, to be considered in decisions

It is essential that standards-based education not be hijacked by improper test use and that tests be used responsibly and in ways that support, rather than undermine, the teaching and learning process.

using test data for purposes such as promotion between grades or determining types of intervention needed;

- professional development for teachers in how to use assessment data to inform instructional decision making;**
- accommodations for special-needs students and individual waivers for special-needs students as determined by their IEP team; and**
- multiple opportunities for students to retake high-stakes tests.**
- support the development of responsible reporting procedures that provide a context for understanding test results by including data, such as class size, numbers of certified teachers, student attendance and mobility rates, that make the results intelligible to parents and the public and useful to teachers and other school personnel;**
- identify and disseminate professional development programs to help teachers use test results and test data for instructional and school improvement; and**
- identify and disseminate professional development programs to help teachers collect and present credible evidence of student learning that goes beyond standardized test scores.**

AFT CONVENTION RESOLUTION

Achieving the Goals of Standards-Based Reform

Adopted July 2002

The Vision

Standards-based education is a relatively new phenomenon in America. In 1983, in response to national concerns that students in America were not learning enough to compete in a global economy and that there was an intolerable gap between the achievement of minority and non-minority students, Albert Shanker, then president of the American Federation of Teachers (AFT), urged states to learn from other high-achieving countries and set high and rigorous standards for all children and do what was necessary to make sure that they all had an opportunity to achieve them.

The AFT recognized that such a reform addressed both equity and excellence concerns facing public education. It could assure that no student was left behind, that all students had a rigorous curriculum and that more could be achieved by students and by schools. Setting high, clear standards for all students makes it much more difficult to sustain a two-tier education system—with one level of education provided to poor children and another level provided to everyone else. Without standards, “substandard” work is almost impossible to define. As a result, states and school districts may fail to acknowledge students’ problems and evade the responsibility for providing the academic supports necessary for high student achievement.

The AFT understands that a complex, coherent and sustained effort at school improvement is necessary to achieve the goals of standards-based reform. States and districts must:

- Develop high-level content and performance standards for what students should know and be able to do.
- Develop curricula aligned with the standards.
- Develop the capacity of schools and teachers to help students meet the high standards.
- Develop new assessments aligned with the new standards and curriculum.
- Develop an incentive and accountability system that uses the results of assessments and other variables to provide intervention to school systems and schools that fail to move their students toward high standards.
- Phase in an incentive and accountability system based, in part, on assessment results.

The State of Standards-Based Reform

While much still needs to be done, America is making progress toward a standards-based system. According to *Making Standards Matter 1999*, the AFT's annual survey of state progress on standards:

- every state in the nation, with the exception of Iowa, has raised its standards for academic content;
- all but three states have committed to measuring student achievement toward the standards;
- 23 states have or are developing incentives (advanced diplomas, free college tuition) to motivate students to achieve at higher standards; and
- 29 states require and fund academic intervention programs for students who are struggling to meet the standards.

Despite this progress, we are still far from “a coherent and sustained effort at school improvement.” In too many instances, states and districts have paid more attention to the development of “gotcha” assessments than to improving the quality of teaching and learning. And states have imposed high-stakes assessments without developing curriculum to support the standards, allocating extra resources and supports to students and schools struggling to meet the standards, and implementing professional development for teachers.

Let us be clear. Assessment is an integral part of standards-based reform. Well-designed tests can provide an objective measure of how well students are doing and provide youngsters with a strong incentive to study hard and do well. However, it is important to remember that in a standards-based system, the primary purpose of assessments is not to sort “winners” from “losers”—it is to ensure that all students have the knowledge and skills that they need to succeed at the next level and to trigger assistance for those who would otherwise fall through the cracks.

The Promise

Research confirms the positive effects of curriculum-based assessments coupled with fair incentives, particularly for poor children. Indeed, several U.S. and international studies show that fair incentives can help to improve instruction, target resources to underachievers, and focus attention on and raise the performance of the lowest achievers. Indeed, despite some of the limitations in current implementation of standards-based reforms, we have begun to see improvement in student achievement:

- According to the National Assessment of Educational Progress (NAEP), for example, reading scores for 9-year-olds in the nation’s highest-poverty schools (those with 75 percent or more students in poverty) increased by almost 8 percentage points between 1992 and 1999—almost one full grade level.
- More students are taking and passing challenging academic courses. Between 1982 and 1994, the proportion of U.S. high school students who took Algebra I increased from about 20 percent to more than 65 percent. At the same time, the proportion of African-American youngsters taking Algebra II rose about 20 percentage points.
- Fewer students are dropping out than in the 1970s and 1980s—the improvement is especially striking for black students; and
- Scores are up on both SAT and ACT exams, as are student test scores in many of our most troubled school districts.

The Problems

Despite public support for standards-based reform and rising achievement rates, there are implementation concerns that need attention now. In too many instances, school officials are pursuing standards as just another top-down reform, divorced from the needs and realities of the classroom. States and districts are paying too little attention to developing the curriculum necessary for achieving the standards. The necessary professional development and supports for children at risk are too often unavailable, and where available, may be of questionable quality. And, tests unrelated to the standards and curriculum are being imposed with consequences to students, teachers and schools.

While concerns about implementation take many forms, a central issue has been the inadequacies of the tests, particularly when they are used for high-stakes accountability. In several regions of the country, a backlash against high-stakes testing has begun to develop. In Michigan and Massachusetts, for example, a few students recently launched boycotts of their high school exams, saying that the tests were unfair measures of achievement. Critics have questioned whether high-stakes tests discriminate against poor and minority students. And, according to the executive director of the Virginia School Boards Association, that state's assessment system was implemented in a manner that was logically backward: "testing first, training teachers second, and purchasing new books and teaching materials third." This situation is not unique to Virginia.

The Solution

If we want every student to reach higher standards, states and school districts must pay more attention to proper implementation. Sufficient resources must be available to get the job done. Teachers must understand what the standards are and how to teach to them. They must be provided professional development that focuses on deep content knowledge, clear instructional strategies, and the assessment tools necessary for determining student progress toward meeting the standards. Tests must be aligned with the standards and curriculum. They must define responsible levels of acceptable performance. And most important, students who are having a hard time meeting the standards must get the help they need—and get it early. Where these essential supports are missing—as they are in too many states—failure rates are excessive

and students and their parents have become frustrated and angry. If these problems persist, the promise of standards-based reform will remain unmet.

Recommendations

Given the current context for the development and implementation of standards-based reforms and high-stakes assessments and accountability systems, in particular, the AFT recommends:

In regard to standards, states should:

- Explain and justify the standards they set and the performance levels they require for meeting them. Parents and teachers rightly ask: “Is the standard realistic?” States should compare their standards, assessment and results with those of high-performing countries and with credible exams in the United States—e.g., Advanced Placement tests, Achieve benchmarks, International Baccalaureate and the like.
- Provide examples of standards and of student work at various grade and performance levels so that teachers, students, parents and the public have a shared understanding of what is expected.

In regard to curricula, states should:

- Involve teachers in the development of curriculum aligned to the standards in the core areas of language arts, mathematics, science and social studies.
- Provide guidance and support to schools so that they attend to and integrate important areas of the curriculum that often are not the focus of standardized testing—e.g., art, music, foreign languages.
- Develop a data bank that includes the standards, curricula materials and exemplary lessons and student work related to instruction in the standards.

In regard to professional development, states should:

- Provide resources to assure that all children, especially those in high-poverty areas, have properly trained and credentialed teachers and that they get the added support and time they need to meet the standards.
- Align professional development to the curriculum and the standards.
- Provide additional funding for the development of quality profession-

al development programs that reflect high academic standards and best classroom practice.

- Insure that professional development is based upon what we know works best—e.g., professional development is embedded in the work-day, takes place mainly in the classroom and school, is peer driven and content focused.
- Work with districts and the teachers union to find the time, resources and effective delivery systems to ensure that all members of the school's instructional staff—including teachers, specialists and classroom paraprofessionals—have access to the supports they need to help all students master more demanding standards.

In regard to assessments and their use, states should:

- Adhere to the principles of a standards-based education system by basing their assessments, especially individual student assessments, directly on the academic standards and corresponding curricula that teachers are expected to teach and students to learn; aligning relevant professional development with the knowledge, skills, and instructional strategies the standards call for; and putting in place timely support systems and other interventions for students who are struggling.
- Phase in the “stakes” related to tests on a timetable that corresponds with implementing the above minimum requirements of a valid, reliable, accurate, and fair standards-based assessment system, thereby ensuring that school systems have provided students the means to learn the knowledge and skills they are accountable for mastering.
- Recognize that testing children below grade three is legitimate and beneficial for diagnostic and instructional purposes only.
- Give students appropriate supports and multiple opportunities before they realize the consequences of high-stakes tests.
- Resist efforts both to make new assessments and/or their pass scores or minimum performance level so low that they defeat the aims of the standards movement or, conversely, to make them so challenging that most students, including highly accomplished ones, will fail to meet standards, which would also jeopardize the aims and accomplishments of the standards movement. Instead, establish benchmarks for different levels of student performance on assessments in a way that raises both the floor and ceiling of achievement for all students.

Continually evaluate the rigor of the tests and their pass scores or performance levels against the goal of continually striving for improved levels of student achievement.

- Work to improve test instruments, including the setting of cut-off scores or performance/achievement levels, to assure that the results reflect students' skills and knowledge at the appropriate grade and performance/achievement levels. To the extent possible, make sample assessment questions and tasks public, along with samples of student work demonstrating various levels of performance/achievement.
- Develop new forms of the test for each administration. This is expensive, but it will serve to reduce the problem of narrowing the curriculum and teaching only to a test.
- Design assessment systems that can yield information about the strengths and weaknesses of students in particular content and skill areas and ensure that this information is provided to teachers, schools and districts in a timely, useful and comprehensible manner so that they may evaluate the strengths and weaknesses of their instructional program, improve professional development and target interventions and resources more effectively.
- Involve teachers and their unions in the development and review of assessment systems and test items or tasks used to assess students against the standards.
- Develop responsible policies for assessing students with disabilities that ensure that they have the necessary accommodations to demonstrate the knowledge and skills appropriate to their development and in support of the goals of their individualized education programs.
- Develop responsible policies for assessing limited-English-speaking students' content knowledge during their transition toward full fluency in English.
- Develop technically sound and practical assessment tools for vocational education students that measure the academic standards in applied instructional contexts.
- Acknowledge and reward student achievement gains, not just absolute levels of academic attainment, including progress necessary to succeed in passing future assessments.

- Report the progress of achievement in schools and districts by categories of student—grade level, racial and ethnic group, socioeconomic status, limited English proficiency, special education students, etc.

In regard to intervention, states should:

- Provide high-quality preschool programs for all students, as well as early intervention for students identified as at risk for meeting the standards.
- Provide adequate resources to ensure that students have access to any extra assistance that they might need to learn the material. This might require smaller classes, alternative settings for disruptive students, extra time with a well-trained instructor—including tutoring, “double-dosing,” before- and after-school classes, weekend classes, summer school, etc.—as well as access to specialists and special programs and services.
- Provide intervention in a timely manner; do not wait until students have failed to meet the achievement levels.
- Help to identify or develop the curricula, materials and instructional approaches that can be used in effective intervention programs.
- Provide the funds for continued implementation and monitoring of such programs.

In Sum

The AFT reaffirms its commitment to high standards for all students and to the development of a standards-based system of education. While the course is correct, the implementation needs refinement. To be successful, we must re-dedicate ourselves to creating a system with high standards, a rigorous curriculum, professional development, adequate resources, an assessment system aligned to the standards and curriculum, intervention for students at risk, and rewards and consequences for achievement.

Questions & Answers

Why not just say ‘no’ to testing?

While tests can be improved considerably, and used more parsimoniously, the AFT does not believe that eliminating testing will be useful or effective in advancing an equity and excellence agenda.

Standardized tests can provide useful data about student learning. When aligned with strong standards and curriculum, for example, test scores can help to diagnose student strengths and weaknesses, and can help identify “holes” or “gaps” in the curriculum and standards. Assessment data can trigger important interventions like providing extra help to struggling students or professional development to teachers.

Student test scores can be and, in fact are being, used inappropriately by some. However, eliminating tests is not the answer—educating test makers, test users, and the public on the appropriate use of tests is.

Why is the AFT calling for a mid-course correction for standards-based reform?

Standards, testing, and accountability are the watchwords of any public figure who speaks responsibly about education today. But the standards reform movement could be in trouble without mid-course corrections. If states do not make the needed corrections, the reauthorization of the Elementary and Secondary Education Act: No Child Left Behind Act, which requires annual testing in grades 3 through 8 and greater accountability for schools, may exacerbate some of the problems concerning testing and accountability.

It takes more than standards and testing to create an effective standards-based education system. Such a system must also include curricula aligned to the standards, professional development for teachers, help for children struggling to meet the standards, and policies that make meeting the standards count. And, states need to develop all of

these components in an ordered and systemic fashion. For example, imposing consequences without also having an aligned curriculum, teacher professional development, and adequate resources is a sure recipe for disaster.

The AFT is also concerned that standards-based education is threatened by a “testing backlash” that has erupted in some states where parents and teachers perceive the tests as excessive and/or unfair. Polls and newspaper articles show that teachers, parents, and other education stakeholders generally support standards-based reform, but they also indicate that these groups have reservations about how it is being implemented, particularly in regard to testing. Specifically, teachers and parents are concerned that testing:

- Narrows the curriculum and eliminates challenging coursework;
- Creates undue pressures on children;
- Encourages cheating;
- Leads to misclassification or exclusion of students;
- Is wrong when a single test is used for high-stakes decision-making;
- Can yield results that are not always accurate; and
- Are disproportionately unfair to minority children.

Whose responsibility is it to develop and implement the various standards-based elements called for by the AFT?

States have the responsibility for ensuring that schools, teachers, and students have the necessary resources and supports to reach the standards.

States, working in concert with education stakeholders including the unions, must develop a blueprint to ensure that they create a standards-based system that is comprehensive and coherent. States must examine the elements of the system—the standards, the curricula, the tests, the professional development for teachers, the supports and incentives for students—as a whole to determine where there are gaps, where the pieces do not fit together, and where more development must be undertaken to assure that the system is educationally sound.

Is the AFT calling for states to develop and mandate curriculum?

The AFT believes that it is necessary to develop a shared understanding of what a curriculum must contain in a standards-based system. We do not support state-mandated intractable “scripted” curricula that provide no flexibility for teachers. We also do not believe that teachers should have to “go it alone.” They should not have to guess at the meaning of standards; they should not have to determine on their own, what the realistic timelines are for instruction. They should not have to sift through scores of documents to find connections to previous and upcoming content.

The curriculum must be sufficiently detailed so as to provide teachers with a better understanding of what the standards mean both in terms of the content to be covered and the level of complexity required in understanding and applying that content. Curriculum should not be a rigid prescription of what and how the teacher is to instruct a class, but, rather, should provide enough examples to allow the teacher, in collaboration with other teachers, to develop a common understanding of the standards. The curriculum must provide detailed guidance to teachers about content, instructional strategies, and student performance levels.

Will the steps called for by the AFT diminish professional autonomy?

These resolutions do not diminish teachers’ professional autonomy. In fact, they enhance professional autonomy. For instance, teachers must be involved in the development of curriculum. We cannot continue to tolerate teachers’ being left to fend for themselves with a list of state standards and no curricula, or any other materials for that matter, that are based on those standards. State standards are not curricula.

There is absolutely no other profession whose practitioners are denied their most basic tools and expected to invent them and try them out, all on their own, while simultaneously practicing their profession. It would be considered intolerable. It is equally intolerable for our teachers and grossly unfair to the children they serve.

How is the AFT addressing teachers' concerns about testing?

When standards, curriculum, and professional development are missing or underdeveloped, we find testing, instead of teaching and learning, driving the system. Teachers and parents are justified in their concerns that there is too much testing; that test results are not always accurate; that the tests are of poor quality; and that the tests are misused—especially when high-stakes decisions are based on a single test result.

Tests cannot and should not drive the education system. They should provide valuable information about whether educational goals are met and about where more or different efforts must be made to assure success. If we had strong academic standards, curriculum, and high-quality standards-based professional development, as this resolution calls for, then the problem of tests driving the systems and becoming the de facto curriculum would substantially disappear.

The AFT calls on states to develop testing programs that are open, fair, and appropriate. The AFT will monitor states' efforts to implement high-quality testing programs. In addition, the national union will produce guidelines for affiliates to use to evaluate their state and local testing programs. The guidelines will include information regarding: indicators of an open testing program; teachers' involvement in the test development and implementation process; professional development for teachers on how to use assessment data to inform instructional decision-making; and the need for strong standards, curriculum, extra help for students, and professional development for teachers to be in place *before* tests are used for high-stakes decisions.

The AFT calls for an “open” testing program.

What does that mean?

Too many parents and teachers believe testing results and test development is shrouded in secrecy. States must pay more attention to proper implementation, particularly to testing policies. An open testing program helps build trust in the assessment system by providing information to parents and the public about the development, purpose, and use of the state testing results.

An open system is not necessarily an aligned system. But it is only with an open system that we can determine if the tests are aligned with the standards. Knowledge about how the test was developed, who was

involved in the development process, how the standards are measured, how the cut-scores are set, and how student achievement data are reported are just some of the critical information in an open system. An open system brings any problems within the testing program to light so that changes can be addressed. It helps assure that state officials will be responsive to the public's concerns about testing and helps build trust in the testing program, which, in turn, helps build acceptance for accountability measures.

How can we be sure tests are used appropriately?

Testing is a crucial part of any responsible standards-based system. Tests provide information on how well the system and/or students are doing and what changes are necessary if we are to improve learning for all children. Schools need to be satisfied that their assessments are tied both to the standards and to the work that is going on in classrooms. And states must be sure that their assessments meet the criteria of the measurement profession in terms of their development, implementation, and use of results. At a minimum, the tests must be valid, reliable, and designed for the purposes for which they are used.

Test results are used to improve instruction, to identify students who need extra help, to target resources to schools or students who are struggling to meet the standards, and to make decisions regarding promotion and high school graduation. Because this is the case, it is important that the tests be based on strong standards. If the standards are vague, or if the tests measure knowledge and skills unrelated to the standards, then the results are suspect in determining how well schools or students are doing in meeting those standards.

Tests should not drive the system. Instead, they should assist in determining whether the system is working effectively. In a standards-based system, the purpose of assessment is not to sort “winners” from “losers,” but rather to ensure that all students have the knowledge and skills they need to succeed at the next level of schooling and to trigger assistance to those students who would otherwise fall through the cracks. The tests must identify students who are having a hard time meeting the standards, and the states must assist districts to ensure that schools have the necessary resources and that struggling students get the help they need—and get it early.

Background Information

Standards-Based Accountability: Ten Suggestions*

Robert L. Linn

History has shown that testing is a popular instrument of accountability and reform for a number of key reasons, including:

1. Tests are relatively inexpensive.

Compared to changes that involve increases in instructional time, reduced class size, training and attracting better teachers, assessment is very low-cost.

2. Testing changes can be implemented relatively quickly.

Other school reforms may take years to implement, and it may take even longer to know if they have improved schooling.

3. Test results are visible and draw media attention.

Poor results in the first year of a new testing program are usually followed by increasing scores in subsequent years, giving the appearance that schools are improving.

Robert L. Linn is co-director of the National Center for Research on Evaluation, Standards, and Student Testing and Distinguished Professor of Education at the University of Colorado at Boulder. He is the current (2002-2003) president of the American Educational Research Association and former chairperson of the National Research Council's Board on Testing and Assessment (1997-2000).

**This article first appeared in 1998 as a policy brief from the National Center for Research on Evaluation, Standards, and Student Testing.*

4. Testing can create other changes that would be difficult to legislate.

Research has shown that state- or district-level testing and assessment requirements motivate changes in curriculum and teaching at the school and classroom levels. It is much more difficult to directly legislate changes in the classroom.

Unfortunately, when tests are used to make major decisions about schools and students, these attractive features frequently result in unexpected problems. Test results may be incomplete or misleading, resulting in poor policy decisions. Nevertheless, the policy need for rapid information about student progress and school quality ensures a continued high interest in educational testing.

A key feature of current school reform efforts is the creation of educational standards, with the federal government encouraging states to develop challenging content and performance standards. Standards-based assessment systems have quickly become a central part of many state reform programs, led by states such as Kentucky and Maryland. Other states, including Colorado and Missouri, are in the midst of implementing their own standards-based assessments. Already we have found that these systems confront the same challenges as earlier assessment programs plus a few new ones. For example:

1. Educational standards at the national, state, and district levels are often inconsistent.

Reviews of state content standards (*Education Week*, 1997; Lerner, 1998; Olson, 1998; Raimi & Braden, 1998) show that state content standards range from very strong to very weak. Different raters oftentimes give different ratings to the same standards, further contributing to the problem.

2. How standards are formulated and measured makes a difference.

The choice of “what” is measured and the quality of the standards and assessments are both important. Table 1 (page 22) reports important differences in student performance in the subjects of geography, history, mathematics, and reading as measured by the National Assessment of Educational Progress, the nation’s report card.

In Table 1, why are only 9% of female students reaching the proficient level in history but 43% reaching the proficient level in reading? While the differences may indeed be differences in performance, it is much more likely that they are due to how the standards were formulated or to the accuracy of the assessments in measuring their respective subjects.

An assessment only in geography would show more males (32%) than females (22%) at the proficient level while the reverse would be true of an assessment only in reading, with males 29% proficient and females 43% proficient. Further, choice of different combinations of the four tests could produce overall results that were nearly equal for males and females or results favoring one group over the other. The choice of what is measured can also alter the apparent differences in performance of racial/ethnic groups or of groups formed on the basis of other characteristics.

Table 1
Performance Differences on NAEP
By Subject and Gender*

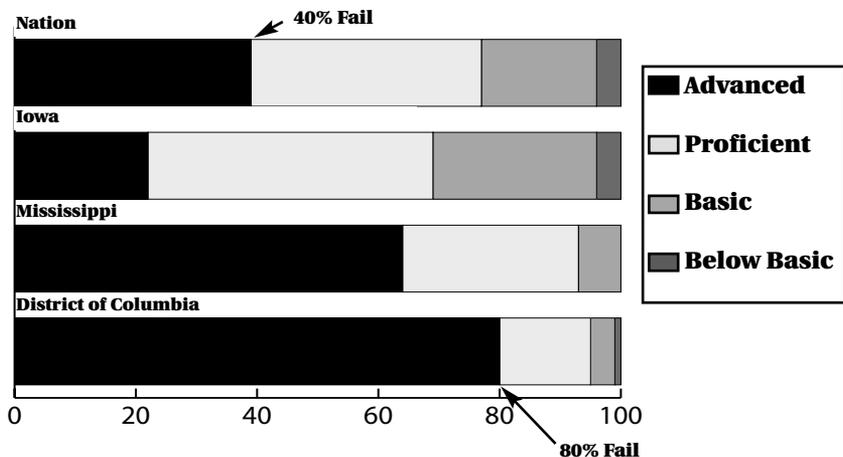
Subject	Males	Females	Difference (M-F)
Geography (1994)	32	22	10
History (1994)	12	9	3
Mathematics (1996)	18	14	4
Reading (1994)	29	43	-14

*Percentage of students at or above the National Assessment of Governing Board Proficient Level. Grade 12.

3. Who's included or excluded in testing can produce different results.

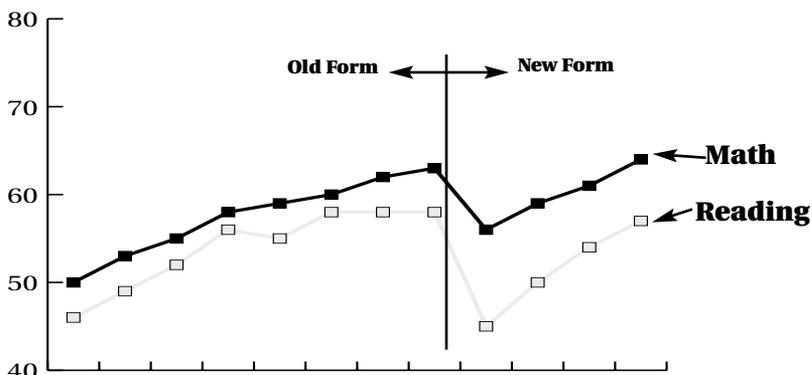
Driven by Title I requirements, standards-based reform emphasizes the inclusion of both special needs students and English language learners in large-scale testing programs. Testing provides important information to policymakers, educators at all levels, and to parents on how all children are doing. However, inclusion can be taken to an extreme. For example, testing students in a language they don't understand will produce inaccurately low test scores. Excluding too many students, on the

Figure 1
NAEP Grade 8 Achievement Levels for the
Nation, Two States and One District*



*Based on Reese et al., 1997.

Figure 2
Results of Changing to a New Test Form*



*Based on Linn, Graue, & Sanders, 1991.

Note that after a period of rising test scores, a new test form is introduced between years 8 and 9. Consequently, test scores drop dramatically in year 9, followed by another steady rise in years 10, 11, and 12. The increase is probably not a result of increased achievement. This is a very typical test score pattern.

other hand, will produce inflated scores. The challenges of meaningful inclusion of all students are difficult, but essential for a credible assessment system.

4. Holding all students to the same high standards will result in unacceptably high retention and failure rates.

Figure 1 shows that nearly 40% of American students did not reach the basic level on the 1996 8th-grade mathematics NAEP test. Are we as a nation prepared to fail or retain as many as 40% of our students nationally or 80% in some districts? To do so would result in major political and legal challenges.

5. Gains in scores do not necessarily signal true improvements.

Research has continually shown that increases in scores on newly implemented tests reflect factors other than increased student achievement. Increases are often a result of teachers teaching to the new test or the use of old test norms (see Figure 2). Standards-based assessments do not have any better ability to correct this problem than other test formats.

6. Different methods may show different student achievement results.

Answers to important questions about student achievement may vary depending on the data analyzed or how it is analyzed and reported. For example, annual testing programs (i.e., fall-to-fall or spring-to-spring) tend to show much smaller achievement increases than testing programs that use a fall-to-spring testing cycle (Linn, Dunbar, Harnisch, & Hastings 1982). The differences may be caused by student selection, scale conversion errors, administration conditions, administration dates compared to test norming dates, practice effects, and teaching to the test.

Ten Suggestions For Policymakers

Despite these problems for standards-based assessment systems, and for most testing in general, there are a number of ways to improve the valid-

ity, credibility, and positive impact of assessment systems while minimizing their negative impact. It is recommended that policymakers:

1. Set standards that are high, but attainable. Unattainable standards lead the public to falsely believe that schools are beyond improvement. Similarly, standards that don't set a high mark will cause the public to lose faith in public schools.

2. Develop standards, then assessments. Studies on the NAEP achievement levels have clearly demonstrated the flaws in attempting to impose achievement levels or performance standards on existing assessments. Revision of existing tests, or creation of new ones, must closely measure the standards and accurately report student achievement.

3. Include all students in testing programs except those with the most severe disabilities. Use accommodated assessments for students who have not yet transitioned into English language programs or whose disabilities require it. This would help to assure accountability for all students and increase the comparability of results for different schools and districts. Report combined scores and separate subgroup scores to provide more accurate and useful information on student and school progress.

4. Useful high-stakes accountability requires new high-quality assessments each year that are comparable to those of previous years. Getting by on the cheap will likely lead to both distorted results, such as inflated scores, and distortions in education, for example, the narrow teaching to the test.

5. Don't put all of the weight on a single test when making important decisions about students and schools (i.e., retention, promotion, probation, rewards). Instead, seek multiple indicators of performance. Include performance assessments and other indicators of success such as attendance, students taking Advanced Placement courses, etc.

6. Place more emphasis on comparisons of performance from year to year than from school to school. This allows for differences in starting points while maintaining an expectation of improvement for all.

7. Set both long- and short-term school goals for all schools to reach. Short-term goals allow for differences in starting positions of different schools. Long-term

goals permit expectations of the same high standards for all by including an expectation that lower achieving schools should have greater annual or biennial growth rates than current higher achieving schools. This combination will give schools a reasonable chance to show improvement, yet help guard against low expectations for schools and students.

8. Like an opinion poll, there is uncertainty in any educational testing system. That **uncertainty should be reported in all test results.**

9. Evaluate not only the hoped-for positive effects of standards-based assessments, **but the unintended negative effects of the testing system.**

10. Narrowing the achievement gap means that we must provide all children with the teachers and resources they need in order to reach our high expectations. This means improving the educational system as a whole, not just more testing or new testing systems.

References

- Education Week. (1997). Quality counts: A report card on the condition of public education in the 50 states. *A Supplement to Education Week*, Vol. 16, January, 22.
- Lerner, L. S. (1998). *State science standards: An appraisal of science standards in 36 states*. Washington, D.C.: Thomas B. Fordham Foundation.
- Linn, R. L., Dunbar, S. B., Harnisch, D. L., & Hastings, C. N. (1982). The validity of the Title I evaluation and reporting system. In E. R. House, S. Mathison, J. Pearsol, & H. Preskill (Eds.), *Evaluation Studies Review Annual* (Vol. 7, pp. 427-442). Beverly Hills, Calif.: Sage Publications.
- Linn, R. L., Graue, M. E., & Sanders, N. M. (1990). Comparing state and district results to national norms: The validity of the claims that "everyone is above average." *Educational Measurement: Issues and Practice*, 9(3), 5-14.
- Olson, L. (1998, April 15). An "A" or a "D": State rankings differ widely. *Education Week*, 17, 1, 18.
- Raimi, R. A., & Braden, L. S. (1998). *State mathematics standards: An appraisal of science standards in 46 states, the District of Columbia, and Japan*. Washington, D.C.: Thomas B. Fordham Foundation.
- Reese, C. M., Miller, K. E., Mazzeo, J., & Dossey, J. A. (1997). *NAEP 1996 mathematics report card for the nation and the states*. Washington, D.C.: National Center for Education Statistics.

Background Reading

AFT Resources

Making Standards Matter 2001

Since 1995, the AFT has tracked states' efforts to implement strong academic standards, align assessments to those standards, provide intervention for struggling students, and hold students accountable for meeting the standards. The 2001 report examines all these issues and reports on states' efforts to develop curriculum resources that help teachers teach the standards (November 2001).

<http://www.aft.org/edissues/pubs>

The Instructional Demands of Standards-Based Reform

If we are to achieve the promise of the standards movement, we need to understand the demands of standards-based instruction and develop the capacity of every classroom teacher to carry it out effectively. Learn about the key differences between standards-based and traditional instructional planning in this paper prepared for the AFT by Kate Jamentz, Director of Professional and Organizational Learning at WestEd (May 2001).

<http://www.aft.org/edissues/pubs>

Principles for Professional Development

The AFT prepared a set of guidelines to help its affiliates review, evaluate, improve, and design professional development programs. Among

the central themes the guidelines highlight are the need for professional development that helps teachers gain deep knowledge of the subjects they teach; reflects current research on teaching and learning; is aligned with the curriculum and standards they use; leads teachers to be intellectually engaged with their colleagues; is job-embedded; and provides teachers sufficient time, support, and resources to master new content and pedagogy (May 2002). <http://www.aft.org/edissues/pubs>

Kissing the Frog: How Teachers Can Transform Testing,

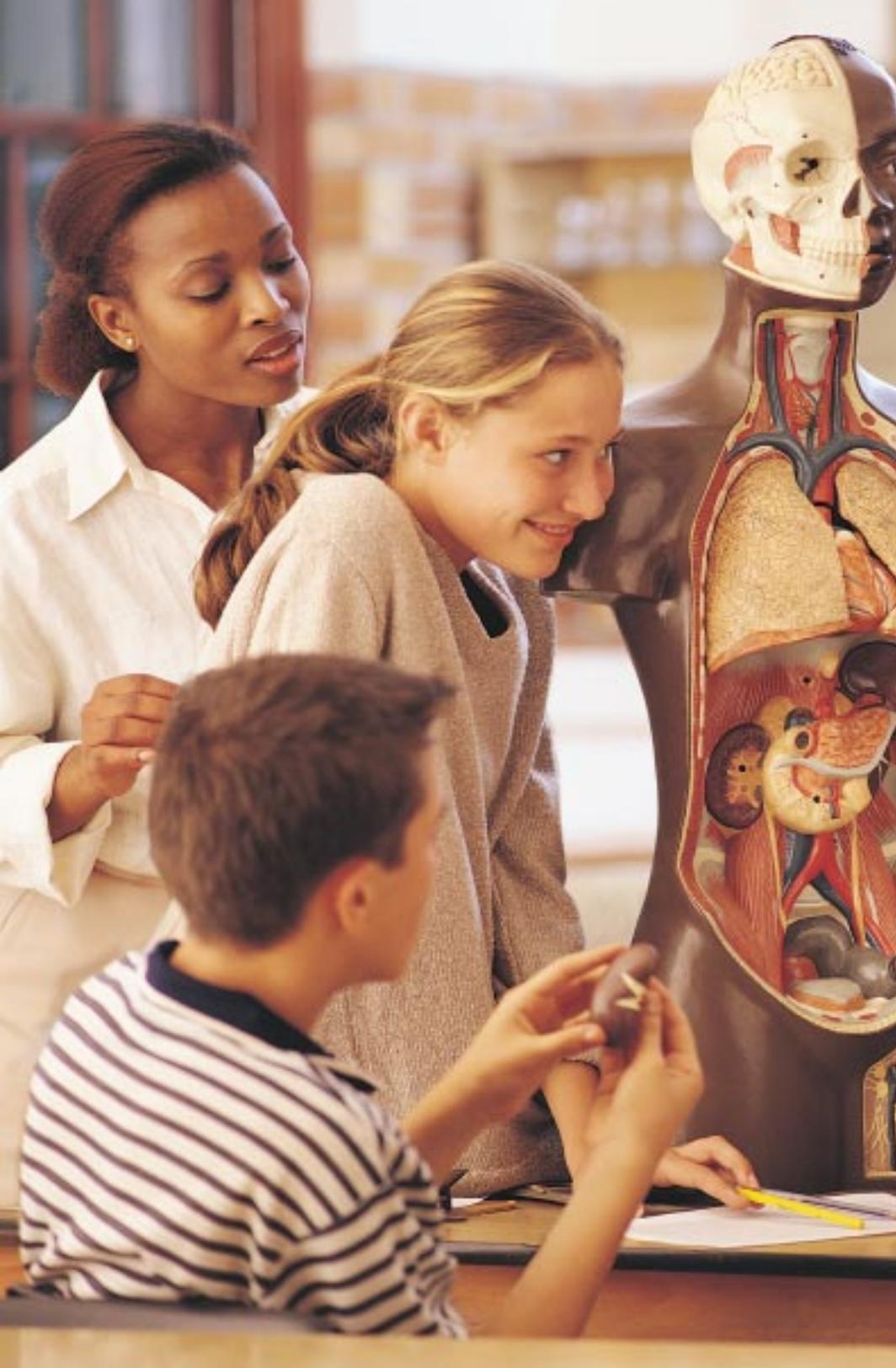
speech by Mari Pearlman at the AFT QuEST Conference, July 14, 2001
<http://www.aft.org/edissues/teacherquality/downloads/Pearlman%20Speech.pdf>

Additional Reading

Thinking About Tests and Testing: A Short Primer in “Assessment Literacy,” by Gerald Bracey for the American Youth Policy Forum, 2000
<http://www.aypf.org/publications/BraceyRep.pdf>

Too Much Testing of the Wrong Kind; Too Little of the Right Kind in K-12 Education, by Paul Barton for Educational Testing Service, March 1999
<http://www.ets.org/research/pic/204928tmt.pdf>

Assessing and Addressing the “Testing Backlash,” by the Business Roundtable, Spring 2001
<http://www.ksagroup.com/ksa/Backlash.pdf>





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

Educational Issues Department
Item Number 39-0246
June 2003

