Needed: Higher Standards for Accountability

Paul E. Barton

While holding schools accountable, NCLB fails to hold itself to basic evaluation standards. Four prescriptions would change that.

During the last quarter century, a model for school standards and accountability has emerged in the United States that is now so locked into state and federal laws that its general shape seems here to stay. If this model is indeed here to stay, we need to keep working to get it right.

The school reform movement, known in its beginning years as standards-based reform, was founded on the proposition that the education system needed to establish rigorous content standards, prepare teachers to teach that content, and align instructional materials and curriculum with those standards. A further proposition was that we could measure a school’s performance in this effort by drawing on existing testing systems, such as norm-referenced tests, that in the past had been used largely to make decisions about individual students. NCLB puts teeth into the practice of using such tests to measure school accountability because it requires states to impose sanctions on schools whose students do not reach designated cut points for “proficient” performance on schedule. NCLB also requires schools to dis-aggregate test results by race, ethnicity, and other factors; and all subgroups of students must meet the standards.

This system clearly looks to standardized tests to hold schools accountable. But the accountability system itself has some accountability problems. I suggest four new directions our school accountability system needs to take.

Apply Basic Evaluation Standards for Using Tests for Accountability

From the beginning, tests used in the school accountability system failed to meet the high standards designed into the system (Barton, 2006). The 1994 amendments to the Elementary and Secondary Education Act that required states to establish content standards and tests gave careful attention to what was needed to produce meaningful test scores in a standards-based reform system. The handbook commissioned by the U.S. Department of Education (Hansche, 1998) told states,

Systems of performance standards and assessments must be created or selected and matched with the content. In an aligned system, all content standards must be accounted for in some manner .... Content standards, performance standards, and assessments must be aligned so that what is taught is tested and what is tested is taught. (p. 21)

According to assessment expert W. James Popham, content standards frequently embrace a
wider range of content for a grade and a subject than a course can realistically cover (Popham, 2004). In this case, aligning a test to a set of content standards is not enough. In such cases, school systems need to prioritize what they plan to teach so that teaching requirements match what is tested. Otherwise, the test will not be sensitive to instruction and test scores will not reflect the results of improved instruction.

Alignment was incomplete in 2001 when NCLB was passed, and it is still incomplete. The American Federation of Teachers (an early supporter of standards and tests) and the Fordham Foundation have evaluated how well all states have met alignment requirements, and their reports agree that there are deficiencies in alignment (American Federation of Teachers, 2006). Other organizations, such as Achieve, have evaluated individual states.

Nevertheless, as soon as NCLB was passed, the Department of Education immediately began using existing test scores to hold schools accountable. Lists of schools “in need of improvement” were compiled, and the sanctions clock started to run.

Measurement experts are explicit about what makes a test “valid” in an accountability system. Where alignment among curriculum, instruction, and assessment is incomplete, the assessment does not meet standards for validity, and we cannot rely on changes in test scores to judge whether schools have become more or less effective. Unfortunately, such scores are being used for sanctions whether the tests meet validity standards or not.

### Hold Schools Accountable for What Goes On in School

Despite all the testing, our present accountability systems do not reliably sort out effective from ineffective schools. Our current methods simply do not measure the change in the knowledge of a student from point A to point B—for example, from the beginning to the end of the school year. Thus they fail to reflect the educational progress over time of any student, class, or school.

So what do our standardized tests now measure? They measure, for example, what students know about a subject at the end of the 8th grade. Our current accountability system compares such scores against a level—or cut point—that someone has judged as “proficient.” Then, it compares these scores with the scores of 8th graders 1 or 5 or 10 years ago. But comparing what certain students know now with what different students knew at the end of past school years tells us little about the quality of instruction. These past students may have known more or less when they entered 8th grade than did the 8th graders more recently tested. Different groups of students will enter 8th grade with a variety of backgrounds in schooling, preparation, and family resources. These differences will have provided each individual with more or fewer advantages for learning. A test score at the end of 8th grade reflects all the learning a student has gleaned from the first 13 years or so of life.

But to hold schools accountable, we need to know how much knowledge students gained in the course of a school year. And we still need to know what students in various groups know at each time we measure: total knowledge at a point in time. The latter data tell us how we are doing, as communities, as states, and as a nation in the entire learning enterprise.

A number of large research studies (Barton, 2006) have measured school progress both ways: student gain during the school year and total knowledge at the end of the year. In each study, the correlation between the two measures was low. This means that some schools whose students are making considerable gains in knowledge in each grade are likely being sanctioned under NCLB; other schools are falling down on measures of student gain but are being let off the hook.

If we had an accountability system that truly measured student gain—sometimes called growth, or value added—we could use this measure to judge whether students in any year have gained enough in that school year to show adequate progress. The end goal should not be achieving set scores by 2014. The goal should be reaching a standard for how much growth we expect during a
school year in any particular subject.

There is much more to a sanctions-based accountability system than just giving students an end-of-year test. There are clear and accepted standards for how to use tests to evaluate the effectiveness of schools. To measure what a school accomplished, you must know what students knew when the school year began and what they knew when the school year ended. Our present accountability system does not operate this way.

Measure Student Gain in a Transparent Fashion

To measure gain effectively during the school year, measurement must be educationally sound; help teachers teach; and be transparent to students, teachers, parents, and policymakers.

Models operating in some schools and research settings do measure student gain, or value added, but none of them meet any of the above criteria. These models use achievement scales developed in one of two ways:

- Tests given at the end of the school year are equated on a scale that shows progress from grade to grade and requires tracking the same students over time. Such tracking, however, may be difficult to do.

- In a second model, test makers create a “stretch” test that covers the subject matter of several grades; progress is determined on the scale of achievement that spans those grades. Using a test covering the subject matter of several grades makes alignment much harder, however.

The models that require achievement scales covering several years are black boxes that teachers and the public can’t see inside. In both models, considerable problems crop up in aligning a test with content taught in an individual grade. Also, both models gauge student gain from the end of one school year to the end of the next school year, even though students’ summer experiences may greatly influence their academic growth. And no test at the end of the school year can help inform a teacher what to teach during the school year.

So is there an assessment method that would give our educational assessment system validity and usefulness? Yes—one that is well-known. It requires creating two forms of a test that measure the same content, with both sections aligned to content standards and instruction—one to be given at the beginning of the school year and one at the end. The resulting gain measures what happened during the year’s instruction. Teachers could use results of the first test to inform their teaching for the academic year and judge their effectiveness for the year by perusing test number two.

True, this method doubles the amount of testing. But a system for judging school quality doesn’t need to test students in every grade every year. Accountability testing could be done on a sample basis, rotating basis, or surprise basis. A lot of testing time could be redirected toward more diagnostic and formative testing: the kind of assessments constructed specifically to help teachers improve instruction. Research shows that such testing raises achievement substantially (Barton, 2005; Black, Harrison, Lee, Marshall, & Wiliam, 2002).

Set Standards for How Much Gain Is Expected

Shifting to an accountability system that measures student gain will go a long way toward improving the system. We then need to set standards for how much gain is expected during a school year.

Although the last year has seen a lot of discussion about measuring growth, gain, or value added, I have heard almost nothing about what should be done with the scores obtained by these methods. After NCLB went into effect, some states claimed that schools in which achievement was “growing” should not be penalized, even though they had not met the yearly target under
the current accountability system. Understandably, the U.S. Department of Education met this claim with skepticism. How much growth would be acceptable?

Just as the current system includes standards for achievement, so must a revised system. We must determine how much growth we expect in 8th grade mathematics, for example. Such judgments must be informed by sound knowledge about how much students in U.S. schools actually learn from year to year. With this knowledge, we can see what amounts represent the low end and high end of actual student gain and systematically set a standard for gain.

This new approach would not let schools off the hook. Our education system could still set high expectations for how much students learn. Standards would apply to all schools equally, including underperforming schools in the well-off suburbs. By continuing to disaggregate student gain scores by race and ethnicity, we could bring into the open what kind of growth is needed to close achievement gaps.

But most important, the spotlight would be on what schools accomplish and what they should accomplish in raising students' knowledge and skills during each school year. Tests will measure this increase, and we could compare test scores to a standard of expectations.

The Need to Persist

Our existing test-based accountability system arrived at its current form in steps, with each step building on what was already there and what could be done at the time. But now we need to take a giant step if we are to have a valid system.

The education assessment system now used cannot be considered valid under ordinary standards of program evaluation because it does not do what it is supposed to do: sort effective schools from ineffective ones. This lack of validity has huge consequences. We should persist until the United States truly has high standards for our schools and for our school accountability system.

References


Paul E. Barton (paulebarton@aol.com) is an education consultant and writer. He is the author of *Unfinished Business: More Measured Approaches in Standards-Based Reform* (ETS Policy Information Center, 2005).