

## The Right Way to Measure Growth

Paul E. Barton

**A defensible test-based accountability system must hold schools accountable—for what happens in schools.**

There is a basic flaw in our current sanctions-based accountability system: its reliance on an end-of-year test. Of course, we didn't design this system from scratch. Rather, it evolved. Starting in the mid-1980s, a standards-based education reform movement emphasized beefed-up subject-matter content, with a test given at the end of the year to see whether students had "gotten" it. Many states, and finally the federal government in No Child Left Behind (NCLB), took these end-of-year scores, compared them with the scores of students from previous years, and used that as the basis for applying sanctions to schools.

However, this type of system does not measure the progress of any student, class, or school over the course of an academic year. End-of-year tests measure students' total knowledge, no matter when or where they acquired it—in preschool years, in the family, during the summer, in previous grades, or in previous schools. Moreover, comparing such end-of-year scores does not take into account how students in one academic year differ in their educational preparation, as a group, from students in another academic year. Many neighborhoods change over time because of flight to the suburbs, gentrification, and immigration.

In any evaluation of program effectiveness, it is widely accepted that evaluators must measure what students already know when the program begins and compare that with a measure of what students know after the program ends. Such "before" and "after" measures are basic.

The current system's flawed approach has been pointed out many times. A number of research projects—using both an end-of-year testing approach and a measure based on student gain during the year—have established a low correlation among the schools identified as ineffective by the two approaches. In other words, a school found ineffective by current measures may nevertheless show gains; conversely, a school showing little gain may be found effective by current measures. The consequences of identifying the wrong schools as ineffective are huge. This misidentification can result in such practices as inadvertently transferring students from effective to ineffective schools (see Barton, 2005).

The U.S. Department of Education has finally begun to see the logic of measuring student gain while students are in school and, as a result, is approving pilot projects for developing and applying this approach, variously called *value added*, *growth*, and *gain*. However, despite the seemingly unanimous judgments rendered by professionals from the evaluation and research communities, no change is contemplated in the law's current end goal of having all students reach the "proficient" level by 2014—a goal set in terms of the flawed approach of using end-of-year scores that gauge the total of student knowledge. If it is the wrong way to meet the target for 2007, it is the wrong way to meet the target for 2014.

### "It Makes Sense, But ... "

I have found that when I talk with people about measuring gain, they don't question the logic of it. They have doubts, however, about switching to that type of measure for school accountability. Some have concerns about how this would play out in the political environment—a reference to the wide acceptance of the current approach. Others fear that using a measure of student gain might slow progress in narrowing achievement gaps. Still others fear that far fewer schools would receive an "in need of improvement" designation, thus reducing pressure on the system to produce better results.

After all, because schools have measured achievement gains within the school year only on a limited basis, we do not have a clear picture of what actual gains might look like. For example, we do not know the variation in gain scores among all 8th grade mathematics students or the differences among average gain scores for different racial or ethnic subgroups or between poor and nonpoor students. If we want to hold schools accountable for the learning that takes place inside them, we need this information, even though it is stepping into the unknown. The way we now measure for school accountability leaves us in the dark: We don't know how much students have learned during the school year, and we don't have the information with which to set standards for how much students should learn.

## Gaps in Gap-Closing Efforts

Without such information, we also don't know how much we need to accelerate achievement gains of minority and poor students to put us on a trajectory to reduce achievement gaps. Because one of NCLB's goals is to close such gaps, we should reflect on the limited way it actually could work. NCLB requires that by 2014, 100 percent of all designated subgroups reach the cut point score that each state has labeled "proficient" on an achievement scale. This cut point varies among the states, and despite its attractive name, "proficient" is a designated minimum that all students are expected to achieve. In that sense, it is like the "minimum competency" movement of the 1970s. How much the gap in scores narrows depends entirely on how high this minimum is set. If it is set low enough, it can have little or no effect on narrowing the gaps in average scores.

Moreover, although the percentage of black students who score at or above this minimum may rise, the scores of white students may also rise. For example, from 1999 to 2004, average math scores of 9-year-old minority students (black and Hispanic combined) rose on the National Assessment of Educational Progress (NAEP). But average scores of white students also rose. Although NCLB has a goal of raising *all* groups toward a common cut point on the scale, scores in the top half of the achievement distribution are also rising. This is a result of a large increase in students taking advanced placement courses, an American Diploma Project working to enable students to earn college credits in high school, and a major push to make the United States competitive in science and mathematics. And before standards-based reform morphed into test-based accountability, it was about beefing up the content of instruction for all students.

A focus on disaggregated *gain* scores would reveal the inequalities in achievement at all levels along the achievement scale. Do we want to settle for equality in the percentage of students reaching a single cut point attractively called "proficient," or do we want to expose inequality in the high-achieving suburbs as well as in low-income inner cities—and all up and down the line?

## Getting It Right

If we are to have a test-based accountability system, it needs to be a defensible one that holds schools accountable *for what happens in schools*. We can create a system that uses gain scores to ensure that schools do their jobs as determined by some rationally set standard of how much students should learn over the nine-month school year. Schools not doing their jobs would run the risk of being sanctioned.

Although several places in the United States (such as Tennessee and Dallas, Texas) have considerable experience with using a value-added approach, there is nothing simple about creating a new system of accountability to measure student gains during the school year. I worry that these systems are black boxes constructed by psychometricians and statisticians who create "vertical scales" and "equate tests" from one year to the next, with perhaps a set of "adjustment factors" for variations in performance that result from student characteristics and not instruction. Such systems will not be transparent, and no one will understand them. A precious few people will be able to tell whether they are valid or not, and nothing in a scale score change from one year to the next will help teachers improve instruction. Systems like these will not be able to factor out important differences in students' summer experiences, although research has documented the reality of a summer loss for some and a summer gain for others.

But one system would be both transparent and understandable—and would deal with the issue of summer loss or gain. It would involve giving students a test aligned with the curriculum in both the fall and the spring. We could use the familiar technology of creating two forms of the same test to measure gain during the year, disaggregating gain scores as we currently do under NCLB. This would inform teachers of each student's deficiencies at the *beginning* of the year, which current end-of-year testing does not do. And it would provide a test that helps teachers teach, although it would not be strictly a "formative" assessment. We would need to set standards for how much gain to expect during the year.

Yes, this system requires twice-a-year testing. But we can have an accountability system that does its job without requiring schools to administer such tests in every subject, grade, and year. We can do the testing every several years and use that extra time in between to engage in formative assessment practices that improve instruction. As Kurt Landgraf, president of Educational Testing Service, argues, "We've got to stop using assessment as a hammer and begin to use it appropriately, as a diagnostic and learning tool."

When we have a system of measuring gain in achievement while students are in school, we cannot mix it in some sort of hybrid approach with the current end-of-year proficiency standard that is based on a student's total knowledge accumulated from birth. However, mixing them is just what the National Governors Association (NGA) proposed in April 2007, in a statement endorsed by the Council of Chief State School Officers and the National Association of School Boards. The National Conference of State Legislatures refused to endorse the NGA recommendations. David Shreve, senior committee director for the National Conference, had it right: "You can't have a growth model that then requires the absolute proficiency performance at the end of the process." (Klein, 2007). To do so defies logic, like trying to mix oil and water.

## A Dual Approach

What we need is a dual approach that would, on one hand, hold schools accountable for meeting standards of student gain and, on the other, provide help to schools with a disproportionate number of students who score low on end-of-year tests. We need to raise the achievement in these schools—whether or not the schools meet accountability standards or are subject to sanctions. Such an approach ensures that we will not let up in our efforts to raise achievement and close gaps in switching from one type of sanctions-based accountability system to another.

When schools fall short of meeting carefully developed standards of how much students should gain while in school, the requirements imposed on them should be evaluated and shown to be effective interventions. Schools with high proportions of low-scoring students (as shown by scores disaggregated by race/ethnicity and income) could undertake a variety of efforts that have proven beneficial. Efforts could include

- Hiring experienced and highly qualified teachers and getting adequate instructional materials into the schools, dealing with disruptive student behavior, and using technology in instruction.
- Tutoring and mentoring individual students to supplement the often meager resources available in their homes, especially in low-income, one-parent homes.
- Extending instruction and enrichment programs beyond the regular school schedule, such as on Saturdays or through longer school days or summer programs, recognizing that some students will not be able to make up deficiencies during the regular school day because the other students aren't standing still.
- Increasing efforts for parent involvement and participation: for example, by giving low-income families books and other resources or providing literacy instruction to enable parents to use those resources.

If we are to close the achievement gap, we must also broaden our education policy to include programs that solve nonschool problems that interfere with learning. For example, we should address health issues like poor eyesight and

provide adequate nutrition beyond the school lunch program. Community schools are now taking on such matters through alliances with other agencies in the community.<sup>1</sup>

If we are to use total knowledge at the end of the year as the criterion for setting high goals for students, then we need to address all life conditions and experiences that result in the total of what students know and can do—for example, circumstances like low birth weight because the mother had poor nutrition or no prenatal care. An ambitious goal of eliminating achievement gaps in total knowledge requires an ambitious effort to deal with the conditions and experiences that contribute to creating those gaps.

## An Unequal Start

This dual approach—which measures both gain *and* performance on end-of-year tests—recognizes that low-scoring schools may or may not meet the gain standards for instruction. Some schools enroll students who start far behind their peers. Although instruction may meet standards, those students do not come near to catching up in a regular school day in a regular school year.

This is exactly the contention of charter school advocates when faced with NAEP analyses showing that students in charter schools do no better or worse than students in regular public schools. Charter schools, their advocates point out, enroll students who enter further behind than students in comparison schools. If you have doubts about how far behind many students are when they start school, examine the evidence from the U.S. Department of Education's Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (West, Denton, & Germino-Hausken, 2000) and Hart and Risley's (1995) ground-breaking study of differentials in the development of vocabulary in the first three years of life, not to mention the large numbers of English language learners who receive little help in their English studies from their parents.

This dual approach would use test-based accountability to identify schools that do not measure up in terms of how much students learn during the regular school day and school year. It would make time for the use of formative assessments and implement proven approaches to raise the achievement of students in all low-scoring schools.

## References

- Barton, P. (2005). *Unfinished business: More measured approaches in standards-based reform*. Princeton, NJ: ETS Policy Information Center.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Brookes.
- Klein, A. (2007, April 6). Governors enter fray over NCLB. *Education Week*.
- West, J., Denton, K., & Germino-Hausken, E. (2000). *America's kindergartners: Findings from the early childhood longitudinal study, kindergarten class of 1998–99, Fall 1998*. Washington, DC: U.S. Department of Education. Available: [www.nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2000070](http://www.nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2000070)

### Endnote

<sup>1</sup> For examples of successful initiatives, visit the Coalition for Community Schools at [www.communityschools.org](http://www.communityschools.org).

*Author's note:* I wish to thank Jack Jennings, W. James Popham, Diane Ravitch, Bella Rosenberg, and Andrew J. Rotherham for their comments, suggestions, and encouragement in preparation of this article.

---

**Paul E. Barton** is an independent writer and consultant and Senior Associate (and former Director) of the ETS Policy Information Center. He is the author of *Unfinished Business: More Measured Approaches in Standards-Based Reform* (ETS

Policy Information Center, 2005) and *"Failing" or "Succeeding" Schools: How Can We Tell?* (American Federation of Teachers, 2006); [paulebarton@aol.com](mailto:paulebarton@aol.com).

---

Copyright © 2007 by Association for Supervision and Curriculum Development