Testing for GROWTH

It makes sense to measure students by their year-to-year progress rather than their test scores.

Gerald W. Bracey

Many people are dissatisfied with how the No Child Left Behind (NCLB) law determines Adequate Yearly Progress (AYP), and not just because it is arbitrary and leads to projections of high failure rates. In its current form, AYP is not really a measure of progress. Instead, it looks at this year’s third graders and compares them with last year’s third graders. To measure genuine progress, we would need to look at what last year’s third graders are doing in the fourth grade in order to measure individual growth.

IN BRIEF

One of the criticisms of the No Child Left Behind Act’s Adequate Yearly Progress (AYP) provision is that it fails to take into account year-to-year gains that fall below the state’s AYP threshold. A solution would be an assessment system that tracks individual growth from grade to grade. The best-known model is the Educational Value Added Assessment System (EVAAS), that makes strong claims about student growth with effective teachers. The author notes EVAAS inconsistencies and describes a more promising model developed by the Northwest Evaluation Association.
Some feel that measuring growth would not only be more appropriate, it would be fairer. As things stand at present, a school with a low starting point whose children are showing large gains might still fall below the state’s threshold for AYP.

In the past, most states have lacked the technical capability to conduct such year-to-year tracking, but improvements in information technology now make such tracking feasible and affordable. In fact, the U.S. Department of Education has $24.8 million for grants to help states develop student tracking systems. However, measuring growth is trickier than it might seem at first blush, and determining who is responsible for growth—or lack of it—is trickier still.

A Model for Measuring Growth

The best-known model currently in use for measuring growth is the Educational Value Added Assessment System (EVAAS), originally called the Tennessee Value Added Assessment System, developed by William Sanders and colleagues at the University of Tennessee. This system makes strong claims about the growth of students with highly effective teachers, and the ability to relate that growth to individual teachers. Sanders’ graphs purport to show that scores soar with three consecutive years of effective teachers, while they plummet with three consecutive years of ineffective teachers.

However, the EVAAS definition of effectiveness leaves something to be desired. Sanders developed the system using items from the Comprehensive Tests of Basic Skills, standardized, norm-referenced achievement tests designed to be insensitive to instruction. How then can it measure instructional effectiveness?

In addition, the EVAAS definition of effectiveness is rather circular. An effective teacher is defined and measured by rising test scores. With that definition, we should not be surprised to see that after three years of effective teaching, students’ test scores will have risen.

Finally, the EVAAS model treats each teacher as a world unto herself or himself, with no impact on other teachers. But what if the school values collaboration, team teaching, interdisciplinary curricula, and student autonomy? Haggai Kupermintz, a critic of EVAAS, presents an example of a science teacher and a math teacher collaborating in a computer-rich environment to improve both math and science achievement. Assuming they succeed, who should get the credit?

“As the technology of testing becomes more potent and as—or if—testing becomes more linked to the cognitive processes, we can expect more interest in testing for growth.”
Attempts to disentangle such complex, interwoven contributions of the science teacher, the math teacher, and the computerized learning environment into isolated and independent “effects” are not only methodologically intractable but conceptually misguided. Teaching and learning are aspects of a synergistic phenomenon whereby dynamic forces interact… (Kupermintz 2003).

Effective vs. Ineffective Teaching

Perhaps the best use of value-added techniques like EVAAS has been to expose another inequity in our education system: low-achieving kids are the most likely to get ineffective teachers. Although Sanders denied that his results showed this, an experiment reported by Kevin Carey (2004) clearly revealed it.

In this study, students at the end of grade 4 were categorized as high, medium, or low achievers. The researchers then looked at how many in each category were assigned effective or ineffective teachers in grades 5–7. The results show that twice as many low-achieving students were assigned to ineffective teachers (81) as were assigned to effective teachers (40). Conversely 77 high-achieving students later studied under effective teachers, while only 30 had to cope with ineffective teachers.

The fact that lower-scoring children get less effective teachers is but a concrete example of a more general problem: If we want to attach test scores of children to evaluations of teachers, then each year the children and teachers must be assigned to classes at random. Of course, children are never assigned to teachers at random and so their test scores reflect a host of other influences, including where their parents choose to live and how much pressure parents put on a school to have their child taught by a particular teacher.

It follows that neither EVAAS nor any other value-added system should be used alone to make personnel decisions. For example, a 1991 study by Koretz et al. found that skills displayed on one test didn’t seem to generalize to another. It discovered that scores fell when districts changed from Test A to Test B, but that over a period of years they climbed back to where they had been. However, when Test A was reintroduced, students scored considerably lower than when it had been the original test. Such test score volatility could render data from value-added models nearly impossible to interpret.

A Different Measure of Growth

The Northwest Evaluation Association (NWEA) has developed a different test that does not attempt to link growth to individual teachers. In the NWEA model, the questions students are asked to answer later in the test depend on how well they answered earlier questions. Students that perform well early in the test get more difficult questions as they continue, while students that don’t score as well early might get less difficult or easier questions.

Until recently, NWEA divided test performance into 10 categories, with different growth projections for each. For example, it found that the highest scoring students typically do not show as much growth in the next year. NWEA claims that, given a score of X, it can specify what skills the student has and might not have, and what skills the student has yet to develop. Of course, any such predictions can fail and NWEA has found that students in some schools do not grow as much as predicted, and that some grow more. They do not yet know how some schools are better at inducing growth than others.

As the technology of testing becomes more potent and as—or if—testing becomes more tightly linked to cognitive processes, we can expect more interest in testing for growth. For now, however, the status of value-added models is indicated in a recent report from the National Association of State Boards of Education (2005), which repeatedly uses the words “preliminary,” “promising” and “potential.”

We will have to wait and see if value-added assessment models live up to their promise.

References


Gerald W. Bracey, an independent education researcher, is an associate of the High/Scope Educational Research Foundation, a fellow of the Educational Policy Studies Laboratory at Arizona State University, and author of the recently published Reading Educational Research: How to Avoid Getting Statistically Snookered (Heinemann 2006). His e-mail address is gbracey1@verizon.net.

WEB RESOURCES

The National Association of State Boards of Education provides a summary of its recent report on value-added assessment.

www.nasbe.org/press_release

SAS offers a complete description of the Educational Value Added Assessment System (EVAAS).

www.sas.com/govedu/edu/evaas.pdf


www2.edtrust.org/EdTrust/Press+Room/value-added.htm

RAND has made available online a 2003 monograph, Evaluating Value-Added Models for Teacher Accountability.