

Article published Jan 5, 2005

Ambiguous yearly progress

Every year schools face an annual day of reckoning. That's when officials find out if their school made Adequate Yearly Progress. AYP is the way the No Child Left Behind law attempts to hold schools accountable for the job they do.

NCLB requires specific annual gains toward its decree that by 2014 every student in every school in America will be academically proficient. That means they'll all score high enough to satisfy the new, higher standards that everybody's been talking about ever since we realized that a lot of American students weren't satisfying the old, allegedly lower standards.

This demand for "success for all students" isn't just a lofty, rhetorical goal. It isn't an ideal we aren't expected to attain. It's a federal statute, and it's bristling with sanctions like funding penalties and losing local control of your school.

It's also a pipe dream. Anybody who's ever been to school, anybody who knows any children or adults, also knows that nothing on this earth is ever going to make everybody succeed, even if you set the standards low, but especially if you set the standards high.

That makes NCLB's statutory mandate an objective we're expected to attain even though we can't possibly attain it.

Each year when the stats are released, schools that make AYP don't question the data because no principal wants to suggest that his school isn't as good as it looks.

Schools that don't make the cut can't question the data because it sounds like sour grapes and complacency.

Just once I'd like to hear someone say, "These numbers are nonsense."

So here goes. My school passed, but our numbers are nonsense. Don't misunderstand. I think we're a decent school. But the fact that we made AYP is a happy statistical accident. Next year we might not be so lucky.

That's not just my opinion. Congress's General Accountability Office reported that a majority of state and district officials sampled nationwide experienced problems with "unreliable student data."

A Brookings Institution study found that "50 to 80 percent of the improvement in a school's average test scores from one year to the next was temporary" and "caused by fluctuations that had nothing to do with long-term changes in learning or productivity." According to a senior Rand analyst, the AYP process doesn't identify "good schools" and "bad schools." Instead, "we're picking out lucky and unlucky schools."

Being wrong 50 to 80 percent of the time isn't bad if you're playing roulette, but it's not so hot if you're rating school districts. Lucky and unlucky aren't words statisticians ordinarily like to hear.

What makes the numbers so meaningless? Scoring the new generation of "standardized" tests has grown increasingly subjective and arbitrary.

A Stanford study found that students with skills at the 50th percentile will receive a score within five points of that only 30 percent of the time on standardized math tests and 42 percent of the time on reading tests. This means that scores are typically accurate less than half the time.

A 2003 National Board on Educational Testing and Public Policy report documented a dramatic escalation of "undetected human error." One Houghton-Mifflin error alone "led to incorrect scores for roughly 250,000 students in six states." The recent history of standardized testing is a chronicle of coast to coast data debacles.

Even if we could design valid tests and then score them reliably, comparing last year's fourth grade scores to this year's fourth grade scores still wouldn't make sense. That's because last year's fourth graders aren't the same people as this year's fourth graders. They are even completely the same as this year's fifth graders.

Most schools are small enough that gaining or losing a few brighter or slower students from year to year radically changes the class average.

You wind up measuring ability differences, not how well the school or its students are doing.

Since comparing two different groups of students can't possibly tell you if a school improved, you're probably wondering why we do it that way. Fortunately, there's an innovative alternative.

Unfortunately, the alternative doesn't work well either. It's called "value added" measurement, and it involves comparing how well each student scores in fourth grade with how well he did in third grade.

At first glance, this sounds somewhat sensible. The trouble is tracking, compiling, and analyzing the annual scores of tens of millions of individual

students gets very "complex, difficult, and controversial," according to the director of the University of Maryland's assessment research center.

Students frequently move between schools and states, which adds multiple variables to the statistical soup.

But even when kids stay put, the tests change each year. Comparing scores from two different grade level tests requires that the two tests be equally difficult.

Last February the Washington-based Education Trust named Tennessee "easily the best example" of the value added system. A week later the Tennessee legislature was considering a bill to abandon the value-added system.

The problem lay in adjusting results to account for the varying difficulty of test items from year to year.

This involved post-test scoring "adjustments" that needed to be made by the system's inventor. School officials became concerned when these adjustments altered the results by 40 percent.

In fairness, adjustments don't just happen in value-added systems. A few years back North Carolina adjusted its "faulty" statewide scores when test designers inadvertently overestimated the difficulty of their test and "set passing scores too low."

Here in Vermont we "revised" our scores when they "seemed higher" than they should have been, which gives you some idea just how arbitrary "standardized" testing can be.

Not to worry, though. Our scores were only off by 20 percent.

At least, we think they were only off by 20 percent.

I expect to be held accountable for the job I do.

But how well my students learn isn't solely dependent on how well I teach them. Please don't hold me accountable for all the variables.

Don't hold me accountable for failing to reach an unattainable goal.

And don't hold me accountable with an assessment system that can't be held accountable for itself.

Peter Berger teaches English at Weathersfield Middle School. Poor Elijah would be pleased to answer letters addressed to him in care of the editor.