

Dow Jones Reprints: This copy is for your personal, non-commercial use only. To order presentation-ready copies for distribution to your colleagues, clients or customers, use the Order Reprints tool at the bottom of any article or visit [www.djreprints.com](http://www.djreprints.com)

See a sample reprint in PDF format.

Order a reprint of this article now

## THE WALL STREET JOURNAL

WSJ.com

THE NUMBERS GUY | AUGUST 21, 2010

# Needs Improvement: Where Teacher Report Cards Fall Short

By CARL BIALIK



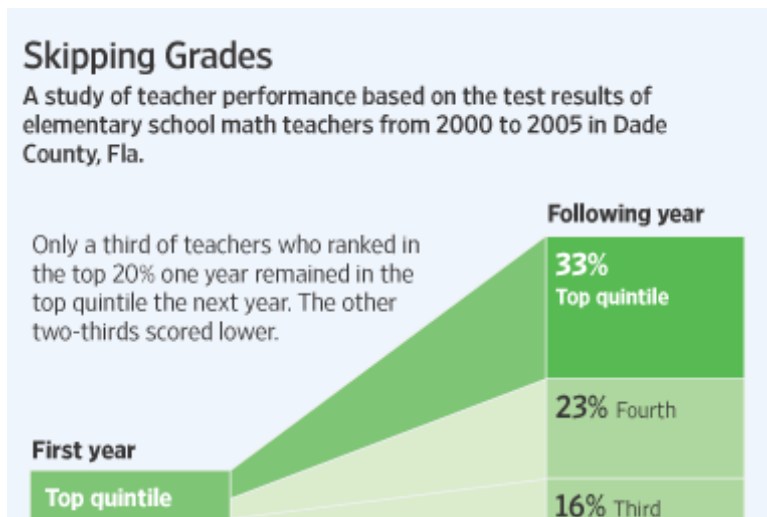
Local school districts have started to grade teachers based on student test scores, but the early results suggest the effort deserves an incomplete.

The new type of teacher evaluations make use of the standardized tests that have become an annual rite for American public-school students. The tests mainly have been used to measure the progress of students and schools, but with some statistical finesse they can be transformed into a lens for identifying which teachers are producing the best test results.

### The Numbers Guy Blog Putting Teachers to the Test

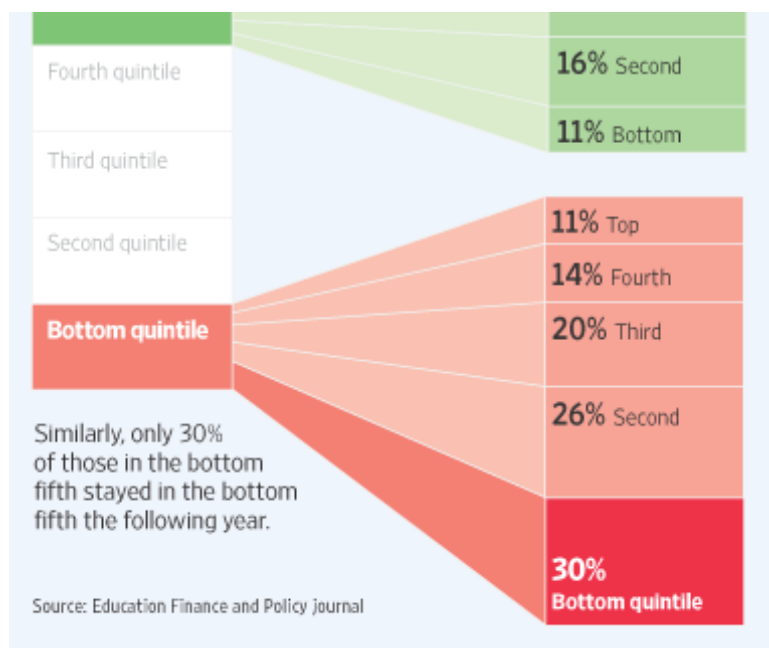
At least, that's the hope among some education experts. But the performance numbers that have emerged from these studies rely on a flawed statistical approach.

One perplexing finding: A large proportion of teachers who rate highly one year fall to the bottom of the charts the next year. For example, in a group of elementary-school math teachers who ranked in the top 20% in five Florida counties early last decade, more than three in five didn't stay in the top quintile the following year, according to a study published last year in the journal *Education Finance and Policy*.



"Because education tends to have this moral-crusade element...we tend to rush to use things before they are refined or really fully baked," says Frederick Hess, director of education policy studies at the American Enterprise Institute, a conservative think tank.

But even skeptics of test-score-based evaluations acknowledge that a uniform, data-based approach for ranking teachers could be superior to subjective methods—such as



principals' observations—that still predominate in schools. "Damn near anything is going to be an improvement on the status quo," says Daniel Willingham, a cognitive psychologist at the University of Virginia.

The U.S. Department of Education has pushed states to loosen restrictions on evaluating teachers through student test scores. To be eligible for a piece of the \$4.35 billion in competitive grants in the Race to the Top federal program, states can't have laws barring a link between student scores and teacher evaluation. And states are scored in part based on whether they evaluate

teachers using test results.

Meanwhile, the District of Columbia began evaluating teachers based on test scores last school year, and fired more than 150 teachers after the school year because of poor performance. Test scores count for 50% of teacher ratings in subjects that are tested.

These measures don't simply ding teachers for their students' low scores, because not all incoming classes start the year equally. Instead, teachers are evaluated based on how much students' scores improve by the end of their year.

But good teachers aren't easy to identify this way. For one thing, students aren't always assigned to teachers randomly. A teacher who gets more than his share of students who learn slowly because of his knack for helping them might be penalized at the end of the year.

## Related

**Taking Schools Into Their Own Hands**  
(08/16/10)

**Obama Defends Teacher Policy** (07/30/10)

**New York City Cracks Down on Tenure**  
(07/30/10)

**Opinion: Giving Lousy Teachers the Boot**  
(07/27/10)

**Teachers Lose Jobs Over Test Scores**  
(07/24/10)

There are other problems with the data. Elementary-school teachers might have just 15 or 20 students in their classes, which is a small sample on which to evaluate a teacher's achievements. "If you're using just one year of information, it's going to be pretty unstable," says Tim R. Sass, an economist at Florida State University.

Research suggests that using multiple years of data helps matters, though only so much. A report from the Department of Education released last month shows

that even with three years of data, one in four teachers is likely to be misclassified because unrelated variables creep in.

Even with these questions, relying on student test scores to create a quantitative assessment of teachers might be better than the current standard practice. At many schools, principals grade teachers based on

a few minutes of classroom observation (and then give most of them high scores). Rating teachers in this way doesn't do all that well in predicting how much their students' test scores will change, according to several studies.

Advocates of the student test-score measure say it can be improved with a carefully constructed model that takes into account such factors as students' family income and schools' support for teachers. Dan Goldhaber, director of the Center on Reinventing Public Education at the University of Washington, points out that some instability in teacher rankings from year to year is to be expected, even desired, as some teachers make more progress than others in any given year.

---

### More Numbers Guy

[Estimates of Religious Populations Require a Bit of Faith](#)

[Online Polling Burnishes Its Image](#)

[Alternative Ways to Gather Census Data](#)

The Los Angeles Times has stirred the debate by commissioning its own analysis of Los Angeles elementary-school teachers. The newspaper published an article about the findings last week and plans to release a database of thousands of teachers' quintile rankings, after giving teachers time to request their rankings and respond for publication.

The Times coverage is helping to raise awareness about the lack of standards for teachers, says Steve Cantrell, senior program officer for the Bill and Melinda Gates Foundation. The foundation is funding a study in seven school districts of teacher evaluation, combining test score-based analysis with other factors, such as teacher tests of subject knowledge and independent ratings of in-class video recordings.

Dr. Cantrell says the research will help determine whether it is possible to create a "persistent and stable measure" of teacher performance that predicts student learning.

**Write to** Carl Bialik at [numbersguy@wsj.com](mailto:numbersguy@wsj.com)

Copyright 2010 Dow Jones & Company, Inc. All Rights Reserved

This copy is for your personal, non-commercial use only. Distribution and use of this material are governed by our [Subscriber Agreement](#) and by copyright law. For non-personal use or to order multiple copies, please contact Dow Jones Reprints at 1-800-843-0008 or visit [www.djreprints.com](http://www.djreprints.com)

