

Veículo: The Wall Street Journal

Data: 24/06/2012

Should Student Test Scores Be Used to Evaluate Teachers?

How much to credit—and blame—teachers for student performance is an issue that continues to confound the education field. To what extent is each student's progress directly attributable to the teacher's efforts? What other factors can determine a student's success? Is there a way to measure each factor separately, including the teacher's influence?

These are just some of the questions that surround the issue of whether student test scores should be used to evaluate teacher performance.

Some say it's unfair to base teacher personnel decisions on student test scores. Students have different levels of ability and commitment, and different experiences outside the classroom. No two students get exactly the same amount of parental support.

Others say that student test scores give an incomplete view but provide a starting point, a basic means of comparison. Combined with reports from trained classroom observers and surveys of how students rate their teachers, supporters say, the test scores may be very useful indeed.

Thomas Kane, a professor of education and economics at the Harvard Graduate School of Education and the faculty director for the Center for Education Policy Research, argues in favor of using test scores in evaluating teachers. Linda Darling-Hammond, the Charles E. Ducommun professor of education and faculty co-director of the Stanford Center for Opportunity Policy in Education, Stanford University, argues against.

Yes: As One of Several Measures

By Thomas Kane

Across the country, school systems are reinventing the ways they evaluate and provide feedback to teachers. Although I don't believe student test scores should be the sole

factor in teacher evaluation, I believe just as strongly that they have an important role to play.

Clear evidence for that conclusion comes from the Bill and Melinda Gates Foundation's Measures of Effective Teaching project, which I lead. The project has been working with 3,000 teacher-volunteers in six school districts to test different forms of feedback for teachers, including their students' gains on test scores. Ours and other recent studies confirm that achievement-gain measures provide valuable information and should not be ignored.

To clarify: We should focus on *gains* in test scores, not end-of-year scores. Any estimate of how much the student has improved while in the teacher's class must take into account the fact that students start at different points. We want to know how much a teacher contributes to student growth during the time students are in that teacher's classroom.

While such student-achievement gains are imperfect measures, the same is true of all measures.

The Teacher's Impact

First, critics say student test scores reflect many things other than the teacher's ability. Other things do matter for a child's achievement gain besides the teacher. But so far, the existing research confirms that individual teachers do have a large impact on student gains.

Harvard University economist Raj Chetty and colleagues studied what happened when teachers with strong or weak records of student-achievement growth either left or joined a school. When the teachers with strong track records left, student achievement in that grade level fell. When they joined a school, achievement rose. (Moreover, achievement remained stable in grades and subjects other than the one where the teacher entered or left.) If the student-achievement gain measures simply reflected the unmeasured traits of students, achievement gains or losses would not have followed the teacher.

Second, despite some fluctuation from year to year, we have found that a teacher's record of promoting achievement remains the strongest single predictor of the achievement gains of their future students. In such a ratings system, a teacher's average

may vary from year to year, but so do the batting averages of professional baseball players. In each case, the measure provides a glimpse (albeit imperfect) of future performance.

Third, although the current state tests focus too heavily on easy-to-measure skills and need to be improved, it's not true that the teachers with larger gains on such tests are simply coaching students for the state tests. In our Gates Foundation study, the students with the largest gains on the state tests also tended to have larger gains on other tests which probed students' conceptual understanding in math and their writing skills. These students also were more likely to report high levels of effort and enjoyment in class.

Moreover, Dr. Chetty and his colleagues found that students whose teachers had high achievement gains on state tests had higher earnings as adults. According to the study, taking aggregate results and comparing classes of similar size, an elementary school class with a teacher in the top 5% of achievement gains is estimated to earn \$250,000 more in the students' lifetimes than a class led by a teacher with average achievement gains.

Other Feedback

Teaching is complex. The right approach to feedback and evaluation is to combine student achievement gains with other measures, such as systematic classroom observations and student surveys. We found that trained observers can identify specific aspects of a teacher's practice that turned out to be associated with greater student achievement.

Moreover, we learned that students as young as fourth grade could reliably identify effective practice, by agreeing or disagreeing with specific statements such as, "In this class, we learn to correct our mistakes," or, "When I turn in homework, I get useful feedback which helps me improve."

No information is perfect. But better information should lead to better decisions. Currently, high-stakes personnel decisions in K-12 education are primarily based on two factors: experience and graduate degrees. In the recent recession, thousands of teachers were terminated based simply on their seniority. As imperfect as the current measures of effective teaching are—and they must be improved—using multiple

measures provides better information about a teacher's effectiveness than seniority or graduate credentials.

A high-quality system of performance feedback for teachers requires money, roughly 2% of teacher payroll costs. Given tight budgets, school systems will have to reallocate resources to cover the cost. But there is no other investment a school leader could make that would offer more bang for the buck.

Dr. Kane is a professor of education and economics at the Harvard Graduate School of Education and the faculty director for the Center for Education Policy Research. He can be reached at reports@wsj.com.

No: Teaching Is Too Complex

By Linda Darling-Hammond

Imagine if your child were graded in the same way that New York City teachers were earlier this year.

The system used to rate the teachers purported to compare teachers' performance against one another. But the scores featured huge margins of error—exceeding 50 percentile points in English language arts and 30 points in math. Thus, if a teacher's rating in English was pegged at the 90th percentile, it might actually have been as low as the 40th, or vice versa.

The ratings were based on students' test scores, analyzed using "value-added" statistical techniques. As in other states, researchers who looked at the data found the ratings were enormously unstable: Teachers who scored low in one year or class were often rated high in another, and vice versa. Teachers working with a large contingent of new English learners or special-education students scored lower than when they taught more-advantaged classes of students. Even teachers of gifted classes were penalized, because their students had already maxed out on the tests.

Clearly, if the scores were measuring a teacher's actual ability, these wild swings would not occur.

What's Really Measured?

Proponents of using test scores concede that such measures are imperfect but argue that they still are useful in the same way batting averages are—as an approximate indicator of performance. But at best, teachers' value-added ratings in one year predict only 25% of the variance in ratings in the next year, leaving 75% or more to be explained by factors such as who is assigned to a teacher's class and what conditions he or she teaches under.

The National Research Council and the Educational Testing Service, among other research organizations, have concluded that ratings of teacher effectiveness based on student test scores are too unreliable—and measure too many things other than the teacher—to be used to make high-stakes decisions. Test-score gains can reflect a student's health, home life and attendance; schools' class sizes and curriculum materials; and the influence of parents, other teachers and tutors. Because these factors are not weighed, individual teachers' scores do not accurately reveal their ability to teach.

Nonetheless, New York City's value-added ratings will soon be used to determine continuation and dismissal of teachers there. And a recently passed state law will extend the practice to all public-school teachers in New York state, not just those teaching reading and math, requiring a dramatic increase in the amount of testing for children.

One-third of the state's principals have signed a letter protesting the new system because they believe it will mismeasure teachers, undermine collaboration and create disincentives for teaching the neediest students. Further, the principals worry that greater focus on teaching to multiple-choice tests will reduce the time for the research, writing and complex problem-solving students need to succeed in today's society.

The Poor Suffer

Proponents of test scores often rightly favor an evaluation method that combines measures of teachers' classroom practice with evidence of student learning, including tests. But for this to work, the test-score measures must be appropriate for the particular students and the curriculum being taught. Unfortunately, federally imposed teacher-evaluation policies insist on using state tests that do not measure growth, are poor measures of higher-order thinking skills and penalize teachers of the neediest students.

Among other things, the tests administered each spring ignore the differences in summer learning between more- and less-advantaged students. More affluent students

have enriched summer experiences, so when they return each autumn, they start school further ahead of where they were in June. Poor students, by contrast, have few opportunities for summer learning. Most actually lose ground between June and September. Value-added measures wrongly attribute this loss to their teachers, further distorting the teacher-evaluation process.

Everyone agrees that teacher evaluation in the U.S. needs to change. But how? Perhaps we should look at what high-achieving nations do. In Singapore, for example, teachers are evaluated by trained observers based on how they support the whole child, from social and emotional development to academic learning; how they strive to improve their practice; and, most important, how they work with other educators to improve practice across the school.

There are also some effective systems in the U.S. that train evaluators to review teachers' instruction based on professional standards, and that look at classroom practice alongside student work. Studies find that frequent feedback from this process increases student achievement, because it helps teachers improve.

Such systems are more complex than test-based evaluation, but they work. For the sake of our children, we should develop thoughtful and accurate measures that support teachers in educating students well.

Dr. Darling-Hammond is Charles E. Ducommun professor of education and faculty co-director of the Stanford Center for Opportunity Policy in Education, Stanford University. She can be reached at reports@wsj.com.