EDUCATION WEEK

Value-Added Evaluation Hurts Teaching

By Linda Darling-Hammond

Here's the hype: New York City's "worst teacher" was recently **singled out and so labeled** by the *New York Post* after the city's education department released value-added test-score ratings to the media for thousands of city teachers, identifying each by name.

The tabloid treatment didn't stop there. Reporters chased down teacher Pascale Mauclair, the subject of the "worst teacher" slam, bombarding her with questions about her lack of skill and commitment. They even went to her father's home and told him his daughter was among the worst teachers in the city.

Now the facts: Mauclair is an experienced and much-admired English-as-a-second-language teacher. She works with new immigrant students who do not yet speak English at one of the city's strongest elementary schools. Her school, PS 11, received an A from the city's rating system and is led by one of the city's most respected principals, Anna Efkarpides, who declares Mauclair an excellent teacher. She adds: "I would put my own children in her class."

Most troubling is that the city released the scores while warning that huge margins of error surround the ratings: more than 30 percentile points in math and more than 50 percentile points in English language arts. Soon these scores will be used in a newly negotiated evaluation system that, as it is designed, will identify most teachers in New York state as less than effective.

Is this what we want to achieve with teacher-evaluation reform?

Everyone agrees that teacher evaluation in the United States needs an overhaul. Although successful systems exist, most districts are not using approaches that help teachers improve or remove those who cannot improve in a timely way. Clearly, we need a change.

As student learning is the primary goal of teaching, it seems like common sense to evaluate teachers based on how much their students gain on state standardized tests. Indeed, many states have adopted this idea in response to federal incentives tied to much-needed funding.

However, previous experience is not promising. Recently evaluated experiments in **Tennessee** and **New York** did not improve achievement when teachers were evaluated and rewarded based on student test scores. In the District of Columbia, contrary to expectations, reading scores on national tests dropped and

> **"As in other professions, good evaluation starts with rigorous, ongoing assessment by experts who review teachers' instruction based**

achievement gaps grew after a new test-based teacher-evaluation system was installed. In Portugal, a **study** of test-based merit pay attributed score declines to the negative effects of teacher competition, leading to less collaboration and sharing of knowledge.

on professional standards."

I was once bullish on the idea of using "value-added methods" for assessing teacher effectiveness. I have since realized that these measures, while valuable for large-scale studies, are seriously flawed for evaluating individual teachers, and that rigorous, ongoing assessment by teaching experts serves everyone better. Indeed, reviews by the **National Research Council**, the **RAND Corp.**, and the **Educational Testing Service** have all concluded that value-added estimates of teacher effectiveness should not be used to make high-stakes decisions about teachers.

Why?

First, test-score gains—even using very fancy value-added models—reflect much more than an individual teacher's effort, including students' health, home life, and school attendance, and schools' class sizes, curriculum materials, and administrative supports, as well as the influence of other teachers, tutors, and specialists. These factors differ widely in rich and poor schools.

Second, teachers' ratings are highly unstable: They differ substantially across classes, tests, and years. Teachers who rank at the bottom one year are more likely to rank above average the following year than to rate poorly again. The same holds true for teachers at the top. If the scores truly measured a teacher's ability, these wild swings would not occur.

Third, teachers who rate highest on the low-level multiple-choice tests currently in use are often not those who raise scores on assessments of more-challenging learning. Pressure to teach to these fill-in-the-bubble tests will further reduce the focus on research, writing, and complex problem-solving, areas where students will need to compete with their peers in high-achieving countries.

But, most importantly, these test scores largely reflect whom a teacher teaches, not how well they teach. In particular, teachers show lower gains when they have large numbers of new English-learners and students with disabilities than when they teach other students. This is true even when statistical methods are used to "control" for student characteristics.

For this reason, Chris Steinhauser, the superintendent in award-winning Long Beach, Calif., where schools have been nationally recognized for progress in closing the achievement gap, refuses to include state test scores in teacher evaluations. He points to one of the district's expert veteran teachers, who routinely takes the highest-need 4th graders. Because she can move such students forward where others often cannot, they gain much more than they otherwise would. Meanwhile, other teachers who have easier classes can experience greater success, and everyone wins.

Penalizing such a teacher for taking on the toughest assignment does not make sense. Rather, Steinhauser has spread this model to other schools, allocating the best talent to the neediest students and supporting teacher collaboration.

"These test scores largely reflect whom a teacher teaches, not how well they teach."

Similarly, Singapore's minister of education explained at last year's International Teaching Summit that his country would never rank teachers by student test scores because doing so would create the wrong incentives and undermine collaboration, which is emphasized in Singapore's schools and teacher-evaluation system. In fact, no country in the world

evaluates its teachers based on annual test-score gains.

Yet this has not stopped some policymakers in the United States from forging ahead. In Houston, where teachers are dismissed or rewarded based substantially on value-added scores, teachers can find little relationship between what they do and how they rate each year. As one put it: "I teach the same way every year. [My] first year got me pats on the back. [My] second year got me kicked in the backside. And for year three, my scores were off the charts. I got a huge bonus. What did I do differently? I have no clue."

Among many teachers recently dismissed was a 10-year veteran who had been voted "teacher of the year." Rated each year as "exceeding expectations," she showed positive value-added scores in most subjects every year, except for the year she taught 4th grade, when English-language learners, or ELLs, are mainstreamed in Houston. The pattern of lower scores in classes with large numbers of ELLs is well known. As another teacher said: "I'm scared I might lose my job if I teach in an [ELL] transition-grade level, because my scores are going to drop, and I'm scared I'm going to get fired." When teachers avoid these classes, high-need students are increasingly taught by less effective novices.

So what's the alternative? As in other professions, good evaluation starts with rigorous, ongoing assessment by experts who review teachers' instruction based on professional standards. Evaluators look at classroom practice, plus evidence of student outcomes from classroom work and school or district assessments. Studies show that feedback from this kind of evaluation improves student achievement, because it helps teachers get better at what they do. Systems that that sponsor peer assistance and review programs also identify poor teachers, provide them intensive help, and effectively remove them if they don't improve.

EDUCATION WEEK

SPOTLIGHT on **Teacher Evaluation**

A collection of essential news features, infographics, and commentaries hand-picked by *Education Week's* editors and available as a downloadable PDF.

If we really want to improve teaching, we should look to such districts for models of effective evaluation, as well as to high-performing countries that have professionalized teaching by ensuring excellent preparation, on-the-job collaboration, and ongoing professional learning.

Linda Darling-Hammond is the Charles E. Ducommun professor of education at Stanford University and co-director of the Stanford Center for Opportunity Policy in Education. Her latest book is The Flat World and Education: How America's Commitment to Equity Will Determine Our Future *(Teachers College Press, 2010).*