

latimes.com/news/opinion/commentary/la-oe-darling-teacher-evaluations-20121105,0,650639.story

# latimes.com

Op-Ed

## A better way to grade teachers

**Effective evaluation requires rigorous, ongoing assessment by experts who review teachers' instruction, looking at classroom practice and evidence of student learning.**

By Linda Darling-Hammond and Edward Haertel

November 5, 2012

It's becoming a familiar story: Great teachers get low scores from "value-added" teacher evaluation models. Newspapers across the country have published accounts of extraordinary teachers whose evaluations, based on their students' state test scores, seem completely out of sync with the reality of their practice. Los Angeles teachers have figured prominently in these reports.

advertisement

Researchers are not surprised by these stories, because dozens of studies have documented the serious flaws in these ratings, which are increasingly used to evaluate teachers' effectiveness. The ratings are based on value-added models such as the L.A. school district's Academic Growth over Time system, which uses complex statistical metrics to try to sort out the effects of student characteristics (such as socioeconomic status) from the effects of teachers on test scores. A study we conducted at Stanford University showed what these teachers are experiencing.

First, we found that value-added models of teacher effectiveness are highly unstable. Teachers' ratings differ substantially from class to class and from year to year, as well as from one test to the next. For example, 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 kind of wild swings hold true for teachers at the top. If the scores were trustworthy measures of a teacher's ability, this would not occur.

Second, teachers' value-added ratings are significantly affected by differences in the students who are assigned to them. Even when statistical models try to control for student-demographic variables, teachers are advantaged or disadvantaged based on the students they teach. Contrary to proponents' claims, these models reward or penalize teachers according to *where* they teach and *what students* they teach, not just *how well* they teach.

We found that a teacher receives a higher value-added score when he is teaching students who are already higher-achieving, more affluent and more versed in English than when he is assigned large numbers of new English learners and students with fewer educational advantages. In fact, when we looked at high school teachers who teach different classes, the student composition of the class was a much stronger predictor of the teacher's value-added score than the teacher. This makes sense: With a classroom full of more-advantaged students, teachers can move faster and cover more material, something the statistical models used for value-added ratings fail to capture.

Finally, value-added ratings cannot disentangle the many home, school and student factors that influence learning gains. These matter more than the individual teacher in explaining changes in scores.

These findings have been replicated in many studies. As a result, most researchers have concluded that value-added scores should not be used in high-stakes evaluations of individual teachers. As the country's leading research organization, the National Research Council, concluded: "VAM estimates of teacher effectiveness ... should not be used to make operational decisions because such estimates are far too unstable to be considered fair or reliable."

What is the alternative? Certainly we need teacher evaluation systems that identify both excellent and struggling teachers based on what they do and how their students learn. And we need systems that help teachers improve, target assistance where needed and remove teachers who cannot, with help, succeed in the classroom.

California's Educator Excellence Task Force recently released a report that outlines the most successful practices internationally. It illustrates that, 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 learning from a range of classroom work that includes (but is not limited to) school or district tests that directly connect with the curriculum and students.

Studies show that feedback from this kind of evaluation improves student achievement because it helps teachers get better at what they do. Systems that sponsor the effective Peer Assistance and Review program also identify poor teachers, provide them intensive help and remove them if they don't improve.

If we really want to improve teaching, we should look to develop such models of effective evaluation rather than pursuing problematic schemes that mis-measure teachers, create disincentives for teaching high-need students, offer no useful feedback on how to improve teaching practice and risk driving some of the best educators out of the profession.

*Linda Darling-Hammond, a professor of education at Stanford University and co-faculty director of the Stanford Center for Opportunity Policy in Education (SCOPE), co-chaired the California Educator Excellence Task Force. Edward Haertel is a professor of education at Stanford University and chairman of the National Research Council's Board on Testing and Assessment.*

Copyright © 2012, [Los Angeles Times](#)