

EDUCATION WEEK

Published Online: December 15, 2010

Public Displays of Teacher Effectiveness

By Rachael Gabriel and Jessica Leste

Recently, a New York state court heard arguments around whether or not to publicly release value-added scores of 12,000 New York City teachers. The court hearing came only a few months after the controversial release of Los Angeles teachers' scores. In late August of 2010, the *Los Angeles Times* began publishing a [series of articles](#) reporting on the quality of teachers and schools in the Los Angeles Unified School District, or LAUSD. In addition to the articles, the paper created a public, online database of individual teacher scores giving readers the power to hunt down "the ineffectives"—those teachers who apparently cause and sustain the achievement gap, many of whom, according to the *Los Angeles Times*, do so unknowingly. In this series, not only were we introduced to "the ineffectives," we were also given one of many tools, value-added measurement, or VAM, to root such teachers out, and more easily identify "the miracle workers"—those teachers capable of leading their students to score higher than a statistical model predicted they would.

When we conducted an analysis of the discourse that surrounds this debate, what was most striking was not the fact that the scores were released, but the ways in which language was used to question and/or silence questions about the implications and outcomes of VAM. Though teacher effectiveness seems like a rallying cry the country can unite behind, the shape of the conversations about its measurement threatens to divide us.

When the *Los Angeles Times* published this series, the newspaper got what it wished for—a nationwide ripple effect—a discourse dispersed in talk and text that simplified and glorified the implications of a useful but not all-powerful tool. Throughout the series, the newspaper set up teachers as either one thing or another: effective or ineffective, good or bad, a detriment or a savior. With McCarthy-era tactics, the paper's series flooded us with profiles of extreme-case formulations—examples so good, bad, or surprising that they almost seduced us into believing that "ineffectiveness" could be lurking anywhere, unbeknownst even to the teacher himself or herself, regardless of certification, reputation, or experience.

The *Times* forgot to share what those who study teacher effectiveness have been arguing for the last decade: Effectiveness is not a monolithic thing, but rather teachers are more or less effective across different subjects, students, and circumstances. So far, conversations about value-added measurement seem to use language in ways that present a single view of teaching and position teacher effectiveness as something static that can be estimated by a single statistic. Those who believe teacher effectiveness is flexible across subjects, students, and varying demands do not

[← Back to Story](#)

Let them play it out
before they live it out

TRY IT FOR FREE! ENTER CODE: edweek

WILL
CAMPUS

Fully Interactive Games that will Improve
Your Teens' Decision Making Skills

suggest that all teachers are good at something—some aren't—but rather that the complexity of roles and expectations for teachers requires them to have a dynamic profile of effectiveness. Those who talk about VAM, as if it were both the crystal ball and the Holy Grail for education reform, would love for us to believe otherwise.

While it may seem that this debate is new news, in 2009, months before VAM was twinkling over the *Los Angeles Times*' presses, several issues of *Educational Researcher*, the pre-eminent education research journal, were devoted to articles that outlined the complexity of identifying, let alone measuring, effectiveness in teaching. Six years earlier, the *Journal of Educational and Behavioral Statistics* published a special issue focusing exclusively on VAM. The overall conclusion of the editors was that VAM was valid only

for school-level, not classroom-level, comparisons. Ironically, concerns around the reliability of value-added measurement are no longer central to the debate about publicly releasing individual teachers' scores. Instead, its validity is most often called into question for the reason summed up by Charles G. Moerdler, a lawyer for the American Federation of Teachers, in [a recent *New York Times* news article](#). "The information has no critical basis other than to facilitate a libel," he said. "If it's garbage in, it's garbage out. Just because it's a number, it doesn't mean it's suddenly objective."

While the outcome of the New York court decision is pending, several New York news outlets, including *The New York Times*, have asked to publish the city's teacher scores. But before New York news sources make the same mistake that their counterparts in Los Angeles made—making VAM seem like a litmus test capable of revealing who is and who isn't an angel or criminal in the classroom—it may be useful to draw upon conversations about VAM that stretch back a bit further than this past August.

Ironically, the very researchers who were popularizing and citing the findings of earlier research aided by value-added analyses are now often quoted in opposition to some of its uses (e.g., Linda Darling-Hammond and Diane Ravitch). Most education researchers are quoted as arguing for "multiple measures" of effectiveness, yet these measures are never described. The plea for multiple measures is therefore constructed as a fuzzy, unknown bundle of "other" things—a soft, teacher-defending, union-loving idea with no evidence, let alone a "real" name. Yet, the names of those "other" things could be readily released: observational data; parent, student, and peer survey responses; portfolio reviews; and lesson analyses. Also, it's important to remember that two 2010 studies, one performed by researchers from Mathematical Policy Research and another by John P. Papay from Harvard University, showed that even if measured twice in the same year, approximately one-third of teachers categorized as "effective" one year were categorized as "ineffective" the very next time, either because effectiveness is subject to dramatic change or because the measure itself is unstable or unreliable to begin with.

In [an op-ed essay for the *New York Post*](#) in October, Joel I. Klein, the outgoing chancellor of the New York City schools, wrote, "So what is value-added data and what can it tell us? It starts with the idea of fairness." His

Though teacher effectiveness seems like a



response, “it starts with the idea of fairness,” makes the concept of VAM seem rational, perhaps even inherently useful. Yet, while Klein purportedly supports the use of VAM, the New York *Daily News* reported that he also acknowledges “that the rating system doesn’t tell the whole story about teacher performance.”

These conflicting perspectives are a construct similar to the logic of the *Los Angeles Times*: Value-added measurement is not perfect, but it’s the best we have. In the end, this not-so-perfect-but-it’s-the-best-we-have approach to measuring teacher effectiveness is positioned as rational, with the questions around the reliability and validity of VAM minimized.

rallying cry the country can unite behind, the shape of the conversations about its measurement threatens to divide us.

The language deployed within this debate is not used to engage in substantive discussion about what is being measured and how. Instead, language is being used to sensationalize the topic, with extreme-case examples often used to counter alternative perspectives. For instance, the lawyer representing the United Federation of Teachers constructed the release of value-added scores as a life-or-death scenario, being quoted in the *Los Angeles Times* as saying: “The city of L.A. did this and a teacher jumped off a bridge. Do we want that?” This not only functions to associate a tragedy with the VAM score release, but also positions those who favor such measurement as supporters of something that threatens the lives of teachers, thus adding urgency to and further polarizing the debate.

Another tactic, which has been taken up by New York news sources, is linking discussions of and references to tenure—a traditionally divisive topic—to discussions of publicly releasing teacher scores. Though peripherally related to notions of accountability, and ways of measuring “effectiveness,” tenure seems to be made relevant almost as a diversionary tactic. This may work to remind people which side they should be for and which they should be against.

We argue that the use or release of value-added-measurement scores does not have to be an issue of tenure, seniority, or job security. Perhaps in New York, there is still a small window of opportunity for a more intelligent conversation—one that puts VAM into context for its readers; one that allows counterarguments to add caution and clarity, not hype and mudslinging, to an already divided and politicized education community. Though it is easy to say everyone is united around the idea of the need for effective teachers to be in every classroom, the exaggerated importance of a single statistic as a means for assessing teachers may divide us once again when it comes to measuring and encouraging the kinds of teaching all students deserve. The way we choose to write and talk about VAM may make all the difference.

Rachael Gabriel is a doctoral candidate in literacy studies at the University of Tennessee, in Knoxville. A former Washington, D.C., middle school teacher, she has focused her research on instruction for struggling readers, the unintended consequences of policy, and the analysis of discourse in educational and political settings. Jessica Lester is a doctoral candidate in applied educational psychology, also at the University of Tennessee. A former special education teacher, she focuses her research on the social, cultural, and historical understandings of disabilities and the application of discourse theory within educational environments.

Vol. 30, Issue 15