



When the Stakes Are High, Can We Rely on Value-Added?

Exploring the Use of Value-Added Models to Inform
Teacher Workforce Decisions

Dan Goldhaber December 2010



When the Stakes Are High, Can We Rely on Value-Added?

Exploring the Use of Value-Added Models to Inform
Teacher Workforce Decisions

Dan Goldhaber December 2010

Introduction and summary

The formula is simple: Highly effective teachers equal student academic success. Yet, the physics of American education is anything but. Thus, the question facing education reformers is how can teacher effectiveness be accurately measured in order to improve the teacher workforce?

There is a growing body of quantitative research showing teaching ability to be the most important school-based factor influencing student performance. The evidence that effective teachers significantly influence student achievement is clear. Unfortunately, improving the effectiveness of the teacher workforce is not a straightforward proposition; while research shows teacher effectiveness to be a highly variable commodity, it also shows that it is not well explained by factors such as experience, degrees, and credentials that are typically used to determine teacher employment eligibility and compensation.

When faced with high-stakes personnel decisions such as laying off teachers, granting tenure, or even paying out bonuses, many school districts, several states, and even the federal government are increasingly pushing for the use of measures of teacher effectiveness. From the Department of Education's Race to the Top initiative that urges states and districts to use teacher performance to inform personnel decisions, to the District of Columbia's IMPACT system that led both to significant bonuses for high-performing teachers and the dismissal of low-performing teachers, educational policy makers and administrators increasingly need transparent and accurate methods to quantify teacher performance.

The importance placed on identifying good teachers and bad teachers stands in stark contrast to the teacher evaluation system. Recent research suggests that teacher evaluation is a broken system. Drive-by classroom visits and binary ratings systems are insensitive to teaching assignments and typically assign unsatisfactory ratings to less than 1 percent of teachers. This "Lake Wobegon effect," where the great majority of a group is characterized as above average, fails to acknowledge

and represent the variation in teacher quality we know exists in the teaching workforce.¹ It is nearly impossible to use many existing evaluation methods for high-stakes personnel decisions such as: When all teachers are above average, how do you decide which teachers to lay off? Which teachers should receive tenure? Which teachers have earned bonuses in a performance-based system?

Given the demand for objective, quantitative measures of teacher performance and the shortcomings of many existing evaluation systems, it is not surprising that a number of districts and states have begun to utilize so called value-added models, or VAMs, to evaluate teachers. Based on the notion that gains in student test scores can be attributed to their teachers, VAMs are designed to measure the impact of individual teachers on student achievement, isolating their contribution to student learning from other factors (such as family background or class size in the early grades) that also influence student achievement.

The use of VAMs is highly controversial and tends to center, at least rhetorically around the notion that VAM measures of teachers will lead to perverse incentives or the misclassification of teachers. I would argue, however, that at least some of the policy debate on this issue masks the more fundamental issue of whether *any* system ought to differentiate teachers and act upon differences.

Today most teacher-evaluation systems rely on observational protocols (by principals or other trained professionals) and generally provide little real information about teacher effectiveness. Part of the reason is that teacher ratings are often on a binary scale where teachers are judged to be either “satisfactory” or “unsatisfactory.” Even when the scale used allows for more nuanced judgments, most teachers receive a top-tier rating that fails to differentiate among teachers to any significant degree.

There are various ways teacher performance measures might be utilized were they to provide more information about the variation in teacher effectiveness. These range from low-stakes uses such as determining professional development, mentoring, or other means of remediating teachers deemed to be underperforming, to high-stakes uses such as compensation, promotion, or lay-off decisions.

When it comes to VAM estimates of performance, we actually know quite a bit. Researchers find that the year-to-year correlations of teacher value-added job performance estimates are in the range of 0.3 to 0.5. These correlations are gener-

ally characterized as modest, but are also comparable to those found in fields like insurance sales or professional baseball where performance is certainly used for high-stakes personnel decisions. Part of the reason that the correlations are only modest is that VAM estimates of effectiveness include measurement error, both because standardized tests are imprecise measures of what students know and because there are random elements such as classroom interaction that influence the performance of a group of students in a classroom.

The fact that measurement error exists may suggest that VAM effect estimates are too unstable to be used for high-stakes purposes because they will lead to teachers being misclassified into the wrong effectiveness categories. This is certainly a valid point to consider, but it is also essential to ground debates over changes to teacher evaluation in what is best for students. Classification error will occur with any evaluation system, but an exclusive focus on the potential downside for teachers ignores the fact that misclassification that allows ineffectiveness in the teacher workforce to go unaddressed is harmful to students. Ultimately, one has to make a judgment call about the risks of misclassification, but it is important to stress here that VAMs should be compared to the human capital systems currently in place and not to a nirvana that does not exist.

The argument for using VAMs is not merely based on the notion that its estimates provide important information about teacher effectiveness, as there is little doubt that they do. Rather it is an argument rooted in the idea that using VAMs is fundamentally important given the evidence that school systems, facing cultural or political constraints, have generally been institutionally incapable of differentiating among teachers. VAMs can be an honest broker when it comes to teacher-performance evaluation, ensuring any performance evaluation system recognizes that teachers are not widgets when it comes to helping students learn. Given this, it should come as no surprise that I believe we ought to experiment with the use of VAM teacher-effectiveness estimates to inform teacher policy.

Concerns about using VAMs are legitimate, but they overlook the fact that any type of teacher-performance evaluation with high-stakes consequences for teachers would be controversial. This controversy, however, rarely arises today because the performance evaluations that are currently being used typically are not high-stakes for teachers, either because they are not designed to be or because the evaluation itself is so inexact that the issue is rarely relevant *for teachers*. But the issue is very relevant *for students*. The misclassifications under the evaluations

governing the teacher workforce today come almost entirely in the form of false positives. I would hazard to say that few would disagree that there is at least some (possibly small) share of the teacher workforce in classrooms who should not be in the classroom despite the fact that they have the credentials and evaluations required to practice.

Unfortunately, much of the policy debate about VAM performance estimates is framed around the potential consequences for teachers rather than focusing on the consequences for students. It is entirely possible that the interests of teachers are not entirely congruent with the interests of students when it comes to teacher evaluation and classification. Certainly imperfect evaluation systems (the only types that exist), for example, that are connected to high-stakes policies, will lead to some incorrect teacher dismissals or rewards. The question however, should not be whether this is good or bad for teachers, but whether the number of incorrect classifications is acceptable given the impact on student learning.

My judgment is that current teacher policies lean too far in the direction of protecting teachers from the downsides of misclassification at the expense of the overall quality of the teacher workforce. It is for this reason that I advocate experimenting with teacher-evaluation system reforms (VAM-based and otherwise) that allow policy to better reflect the variation in performance that we know exists in the teacher workforce.

Given the high-stakes issues of student classroom achievement and teacher outcomes even up to dismissal, it is imperative that teacher evaluation methods provide spot-on performance assessments. The key then is having a system like VAM that truly differentiates among teachers while avoiding the pitfalls of misclassification. Still, regardless of the method used to evaluate teacher performance at the very least it must:

- Be rigorous and substantive while allowing for nuance.
- Provide meaningful teacher feedback.
- Be directly linked to consequences and outcomes.
- Be seen as trustworthy.
- Ultimately result in improved learning and achievement for students.

VAMs can be the honest broker when it comes to teacher-performance evaluation, ensuring that any performance evaluation system recognizes that teachers aren't widgets when it comes to helping students learn. Yet, having said that, VAM is often treated as if it is the magical elixir for all that ails the teacher workforce. There are good reasons to believe this is not the case. Thus, I also recommend that school systems implement a performance evaluation infrastructure that builds confidence in performance measures and provides teachers with timely feedback.

The Center for American Progress is a nonpartisan research and educational institute dedicated to promoting a strong, just and free America that ensures opportunity for all. We believe that Americans are bound together by a common commitment to these values and we aspire to ensure that our national policies reflect these values. We work to find progressive and pragmatic solutions to significant domestic and international problems and develop policy proposals that foster a government that is “of the people, by the people, and for the people.”

