The Value-Added Model and its Appropriate Place in Evaluating Teachers

By: Brian Nussbaum

Education Law & Policy

May 16, 2012
Introduction

The focus and concern on the performance of teachers is at an all-time high in the United States of America. For years, American students set the bar when it came to their performances on standardized tests for math and reading. A great deal of worry about the future occurred when American students began to relinquish that dominance to students from other countries. In many instances, this worry turned to anger from both parents and policymakers, which was focused on the manner in which teachers in the education system were being evaluated. The system seemed to allow for incompetent, or at the very least inadequate, teachers to continue to keep their jobs without any serious concern of consequences for poor teaching. Additionally, teachers who were going above and beyond the call of duty were not rewarded accordingly. As time passed, it became clearer that there needed to be changes in regards to the performance assessment measures of all teachers.

There has been great debate over the past few years regarding the proper way to evaluate teacher performance. This paper will focus on how teachers are currently being evaluated and the appropriate place for the value-added model in making the determination on teacher’s success. This paper will begin by laying the backdrop for the purposes of teacher evaluations in a school district. Furthermore, it will address what the value-added model is and how it is implemented. We will then go into some detail regarding the shortcomings and critiques of the value-added model as a statistic used for summative evaluations. Finally, in light of those critiques, this paper will attempt to show the vast advantages the value-added model can have as a statistic that is used in the formative evaluations of a teacher.
Purposes of teacher evaluations

There are two main types of teacher evaluations that serve very different purposes in the analysis of a teacher’s performance. The two types of evaluations are summative and formative. Summative evaluations are used to make a final decision on factors such as salary, tenure, personal assignments, transfers, or dismissals.\(^1\) The best way to look at summative evaluations is to consider it in the light of a year-end job review. These are often the most stressful types of evaluations for the teaching community. It is this type of evaluation that determines whether they enter the safe haven, commonly known as tenure. Alternatively, formative evaluations are meant to provide teachers with feedback on how to improve performance and what types of professional developmental opportunities will enhance their craft.\(^2\) This type of informative evaluation is grossly lacking within many school districts in America. If a formative evaluation does not take place, teachers who truly want to improve their performance are left without any true understanding of areas in which they can get better.

Although these two types of evaluations are fundamentally different, the best school districts assess a teacher by using measures to give feedback on both formative and summative fronts. Admittedly, simply performing different types of teacher evaluations accomplishes very little if they have no bite to them. For example, a teacher who receives a summative evaluation that purports underperformance and yet lacks any negative consequences for that underperformance will yield an ability for teacher complacency. Similarly, a formative evaluation that simply focuses on how great a teacher is doing, does not give teachers an

\(^1\) “Improving Instruction Through Effecting Teaching Evaluation;,” is a brief written by the National Comprehensive Center for Teacher Quality. This brief was written in February 2008 and was authored by Carrie Mathers, Michelle Olivia, & Dr. Sabrina W. M. Laine. http://www.tqsource.org/publications/February2008Brief.pdf.

\(^2\) Id.
understanding or ability for improvement. With that in mind, it is critical to have measures that will allow for evaluations to have some real substance. If teachers are going to be held accountable they need to be given hard evidence as to how effective their teaching methods have been. If performed properly, these two types of teacher evaluations can be critical to the improvement of teacher’s performance and thus student achievement.

**The Value-Added Model**

The value-added model is a statistical measure that gives teachers evidence as to how effective their teaching methods have been on student achievement. The most recent and mathematically sophisticated version of value-added model was developed by Dr. William Sanders at the University of Tennessee.\(^3\) The goal of the value-added model is to determine the contribution a particular teacher has to their student’s achievement and then compare that level of contribution to those of other teachers. This gives evaluators an ability to show teachers how they stack up amongst their peers. One of the biggest advantages, is the value-added model is objective in nature. Many of the evaluation techniques currently used are subjective in nature and require an evaluator to make recommendations to a teacher simply on observations. The value-added model, in theory, removes a significant portion of subjectivity in a teacher’s evaluation and allows for an objective statistical analysis of performance.

The key to the value-added model, along with its shortcomings, is the manner in which the teachers are measured. The basis of the statistic is the use of past student test scores to predict their performance on future test scores. Once that predictive test score is determined, it is compared to what the student actually received on the test. The difference between the predicted

---

\(^3\) Dr. William Saunders is currently a senior research fellow at the University of North Carolina. He developed the Tennessee Value Added Assessment System that many states have since adopted as a method to evaluate teacher performance.
score and the actual score is then attributed to the teacher as a method in determining the effect they had (positive or negative) on their student’s achievement. In order to increase reliability, the teacher’s score (the difference between the student’s predicted and actual test score) is usually aggregated over a three year period. Once the teacher’s score is determined, it can be compared to the scores of other teachers either within the school district or on a state and national level.

The value-added model is an objective statistic that so many critics of teacher performances consider a necessity in evaluating teachers. It does a relatively good job of accounting for differences in students such as race and gender but, it is far from perfect. Overwhelmingly, this statistic is used by school districts and teacher evaluators in the context of summative evaluations. The value-added number makes up zero to fifty percent of a teacher’s total evaluation in any given state. In many states where this statistic is used, if a teacher has a substantially lower value-added number than the average teacher for multiple years in a row, they will be fired. The fact that this statistic is so significant in determining their job performance causes a great deal of concern for teachers. Additionally, it has led teachers’ unions to attack the value-added model on many different grounds, particularly that of reliability and validity.

There are some clear shortcomings to the value-added model, many of which will be discussed in the next section of this paper. However, the fact that shortcomings exist in a

---

4 “Measuring Teaching Using Value-Added Modeling: The Imperfect Panacea.” This article was written in 2011 by Jimmy Scherrer for the NASSP Bulletin. pp. 122–140.
5 Id.
6 “THE LEGAL AND POLICY IMPLICATIONS OF VALUE-ADDED TEACHER ASSESSMENT POLICIES.” This article examines a few states that use the value-added model and the difficulties associated with using the model. This article was authored by Preston C. Green III; Bruce D. Baker; Joseph Oluwole. Brigham Young University Education and Law Journal. 2012.
statistic should come as no surprise. For the most part, any statistic should be taken with a proverbial grain of salt. The difference is many of those other statistics are not being used to decide whether someone gets tenured or gets to keep their job. The use of the value-added model statistic in summative evaluations is the source for a great number of these issues. Although I do believe there is a place for this statistic in summative evaluations, perhaps the focus on the value-added model should shift to the ever-lacking formative evaluations.

Problems with the value-added model

This section will examine the issues of using the value-added model, particularly in the summative evaluation context. The first problem and perhaps the biggest one, is known as model prediction error. In this case, model prediction error means the probability that based on a certain number of years of data, the model will find that a truly average teacher performed significantly worse than average. A study funded by the U.S. Department of Education, found there is about a 25% chance if using three years of data or a 35% chance if using one year of data that a teacher who is “average” would be identified as “significantly worse than average” and potentially be fired. This is a technical problem in the way that the value-added number is constructed. If school districts are going to use this statistic for summative evaluations, it is rather alarming that someone could be fired for a statistic that has a 25% chance of being incorrect. This is a pretty significant margin of error when one considers that someone’s job potentially hangs in the balance.

---

7 THE LEGAL AND POLICY IMPLICATIONS OF VALUE-ADDED TEACHER ASSESSMENT POLICIES.” This article examines a few states that use the value-added model and the difficulties associated with using the model. This article was authored by Preston C. Green III; Bruce D. Baker; Joseph Oluwole. Brigham Young University Education and Law Journal. 2012.
8 Id.
Similarly, one thing that affects validity is the inability of the value-added model to account for a lack of randomizing students among teachers. Simply put, certain teachers work better with certain types of students and schools try to account for that when assigning students to teachers. If a teacher tends to work really well with a group of students who have behavioral issues, that teacher will most likely get assigned a class that consists of a lot of troubled students. Although this method is in the best interest of the student, it does not help the teacher in terms of their value-added score. Even if the student makes strides in the behavioral aspect of his education, that student’s potential lack of educational achievement will damage the teacher’s value-added score. Alternatively, a teacher whose class consists of mostly gifted students will be rewarded for the academic achievement of their students. This is true even if they did not have as great an impact as the value-added score might indicate. The inability to account for the lack of randomization of students to teachers makes the value-added measure a dangerous statistic to use in a summative evaluation.

Furthermore, a problem arises with the value-added model based on the type of test that is being administered to the student. In Houston, two different types of standardized tests are used to measure a student’s academic achievement. Among those students who ranked in the top category on one of the reading tests, more than 17 percent ranked among the lowest two categories on the other test. Similarly, more than 15 percent of the lowest value-added teachers on one of the exams were in the highest two categories on the other test. This makes the determination of whether a teacher is doing an adequate job very difficult. It is a gross injustice to the teachers, to examine only the value-added score that places them as inferior among their

---

10 Id.
peers. However, you cannot completely ignore the fact that under one of the standardized tests their value-added score is simply unacceptable. This dichotomy causes problems not only for the evaluators in making judgment calls on teacher performance but, also for the teachers, some of whose jobs are at stake based on that number. If jobs are going to be decided based on the value-added model there should not be such vast differences based on the type of test that is being administered to the students.

Moreover, there is a potential for over/under valuing a teacher’s impact on a particular student’s academic achievement. The way the value-added model is currently structured gives the total credit for an increase/decrease in a student’s test score to the teacher who is teaching the student that subject. Regardless of whether a student’s score increases or decreases, there may be numerous factors that can account for the increase or decrease that have absolutely nothing to do with the performance of a teacher.\(^1\) In the case of a decrease, problems at home may lead a child to struggle on the test. Even if a teacher goes out of their way to try and help the student and prevents them from doing even worse on the test, that effort will go unnoticed if the value-added model is the determinative factor in the teacher’s performance. In contrast, if a student’s score increases, that increase could potentially be attributed to a tutor or another teacher. For example, a teacher who is doing a poor job may force a student to get some outside help via a tutor. That tutor then teaches the student the subject matter and that student excels on the standardized test. The fact that the tutor did all the work is lost on the value-added model and credits a teacher who in fact did a poor job with this particular student. The case is even more extreme if another teacher at the school helps the student because that teacher (who is subject to

\(^1\) “Evaluating Value-Added Models for Teacher Accountability.” This article was authored by DANIEL F. McCAFFREY, J.R. LOCKWOOD, DANIEL M. KORETZ, & LAURA S. HAMILTON. It is a comprehensive look at the value-added model and is critical of its effectiveness in gauging a teacher’s success in the classroom. http://www.rand.org/pubs/monographs/2004/RAND_MG158.pdf
the value-added score) will not receive any credit for the additional help he gave to someone else’s student. This aspect of the value-added measure is very difficult to remove but, it is important to realize the potential problem when evaluations are taking place.

Finally, for the most part, the value-added measure is only applicable to those who teach math and science. Only about 40% of teachers would be subject to any type of value-added measure. This would mean that some teachers can be evaluated, at least in part, using an objective test, while others will be evaluated on simply a subjective basis. Additionally, there may be some unintended consequences of only having some teachers evaluated via the added-value model. For example, there may be increased tensions between non-value-added rated teachers wishing to pull students of value-added-rated teachers out of class for special projects or activities. Those teachers who are not subject to the value-added measure might not feel it is a big deal to have a child finish a project and miss some other class time. However, those whose jobs are dependent on the value-added measure are going to want the students in class as much as possible. One of the goals in a school is to create as much sense of unity as possible, especially among faculty and staff. This type of divide may break up that sense of unity and create a negative externality for the overall learning environment of the students.

---

12 “Refocusing the Debate: Assessing the Purposes and Tools of Teacher Evaluation.” This article was authored by John P. Papay. It assesses the reasons for teacher evaluations and some of the available methods currently in use by schools. Harvard Educational Review Volume 82 Issue 1. April 1, 2012.

13 “THE LEGAL AND POLICY IMPLICATIONS OF VALUE-ADDED TEACHER ASSESSMENT POLICIES.” This article examines a few states that use the value-added model and the difficulties associated with using the model. This article was authored by Preston C. Green III; Bruce D. Baker; Joseph Oluwole. Brigham Young University Education and Law Journal. 2012
Advantages of the value-added model as a formative evaluation tool

Formative teacher evaluations are lacking in American education today. One would think that a job as critical as teaching a country’s youth would incorporate a great deal of on-the-job training and feedback. However, there are numerous reasons why formative evaluations do not occur as frequently as they should. It takes a great deal of time to give a truly valuable formative evaluation. With all of the budgetary constraints that schools, especially public ones, face they simply cannot afford to take the time to perform these evaluations or send teachers to receive additional training. This does not mean that formative evaluations should not take place. The value-added model gives school districts a perfect place to begin a formative evaluation. Many of the problems that the value-added model has are reduced or eliminated if it is used in a formative evaluation context, rather than a summative one.

The problem of model prediction error can be mitigated when the value-added statistic is used in a formative evaluation. The statistic itself will still mislabel 25% of “average” teachers as below average. It is not disputed that there are a number of variables outside the control of the teacher that affect the value-added number. There will be too many overreactions on the part of the school district, about a teacher’s abilities if the value-added number is used in a summative evaluation. In the formative evaluation, school districts will be able to examine the figures and determine whether a particular teacher’s value-added number is an anomaly or a trend. This examination will allow for honest dialogue between a teacher and the school district because they do not have the imminent concern of being fired. The hardened stance of a teacher

---

14 THE LEGAL AND POLICY IMPLICATIONS OF VALUE-ADDED TEACHER ASSESSMENT POLICIES.” This article examines a few states that use the value-added model and the difficulties associated with using the model. This article was authored by Preston C. Green III; Bruce D. Baker; Joseph Oluwole. Brigham Young University Education and Law Journal. 2012
being fired after 2 years of a dissatisfactory value-added number is far too harsh. Accountability is crucial in creating good teachers but, rushing to judgment about a teacher because of a statistic that is not foolproof is just bad policy.

Furthermore, the lack of randomizing students to teachers is completely eliminated as an issue, when using the value-based model for formative evaluations. A teacher who works with a majority of students that are troubled will be able to easily explain why some of their students struggle. A school district will be able to judge the teacher’s value-added number through the appropriate lens. It will also limit a teacher who is working with mostly gifted students from resting on their laurels. If the value-added number was just being used for summative evaluations, a professor teaching gifted students would be able to point to their high number and say look how well my students are doing. However, if the statistic is being used to determine how the teacher can improve, there can still be a substantive conversation about whether the students are performing to their maximum capabilities in that teacher’s classroom.

Similarly, if multiple standardized tests are used to evaluate the children the differences in value-added scores for teachers can be used as a developmental assessment. The difficulty caused by having multiple value-added scores, in a summative evaluation, is the school district does not know which one to consider valid. In the formative context, that should not be the consideration at all. Instead of asking which value-added number they should accept as accurate, the school district can determine with the teacher why the value-added numbers are so different. In other words, the school district and teacher can focus on coming up with new ways to teach the subject, so the discrepancy in numbers are not as wide. It also gives teachers the ability to see that their students are struggling with certain types of questions. They can then strive to make improvements in that area through a utilization of new teaching techniques. A teacher that
simply get fired if their scores are too low, will never have a significant opportunity to work with the school district and improve. A formative evaluation allows for the lines of honest communication to be open between a teacher and the school. This can only be considered a positive step for the education of the students.

Lastly, it will be easier for a school district to associate how much credit or blame a teacher should get for a student’s academic achievement. The formative evaluation allows the school district to look at an objective statistic (the value-added model), in a subjective manner, through conversations with the faculty and observations in the classroom. It greatly reduces the importance of what the teacher’s actual number is, without eliminating it, and allows for other factors to be taken into account in determining how that number came to be. A teacher who is tutoring a student after class can be attributed some additional credit. Alternatively, if a school district hears or sees a teacher lecture poorly and yet they still have a good value-based number, these observations can be taken into account. The problem with the summative approach is you cannot change the number once it is on the paper. Due to a policy, a school district may end up having to fire a good teacher and keep a poor one. Moreover, if a teacher’s subject does not include a value-based number then it is simply not talked about in their evaluation. The value-added statistic used as a formative evaluation tool prevents any potential divide between the teachers because their jobs are not determinative on it.

Conclusion

The value-based model is a positive step in the direction of improving teacher evaluations. It is important to incorporate some aspect of student achievement into any evaluation of a teacher. The advantages of the value-based model are clear when used in the
context of a formative evaluation. I will not go so far as to say, the value-based model has no place in a summative evaluation. However, the limitations of the value-based model are exacerbated when performing that type of an evaluation. Therefore, when using the value-based model as a tool for a summative evaluation, I would caution school districts to not place too great a weight on their results. To the contrary, a great deal of a formative evaluation should focus on the factors that encompass the value-based model. It will allow school districts and teachers to speak honestly about ways to improve teaching techniques. Teachers will be more aware of deficiencies and have an opportunity to make improvements in those areas. Potential limitations from the data will be minimized because of the subjective nature in which a school district will be able to analyze the data. Overall, the value-based model is an effective tool when used for formative evaluations of teachers’ performances.