Value-Added Modeling for Teacher Effectiveness

Erin D. Lomax
Specialist in Education Policy

Jeffrey J. Kuenzi
Specialist in Education Policy

December 11, 2012

CRS Report for Congress
Prepared for Members and Committees of Congress

Summary

Two of the major goals of the Elementary and Secondary Education Act (ESEA), as amended by the No Child Left Behind Act of 2001 (P.L. 107-110; NCLB), are to improve the quality of K-12 teaching and raise the academic achievement of students who fail to meet grade-level proficiency standards. In setting these goals, Congress recognized that reaching the second goal depends greatly on meeting the first; that is, quality teaching is critical to student success. Thus, NCLB established new standards for teacher qualifications and required that all courses in "core academic subjects" be taught by a *highly qualified teacher* by the end of the 2005-2006 school year.

During implementation, the NCLB highly qualified teacher requirement came to be seen as setting minimum qualifications for entry into the profession and was criticized by some for establishing standards so low that nearly every teacher met the requirement. Meanwhile, policy makers have grown increasingly interested in the output of teachers' work; that is, their performance in the classroom and the effectiveness of their instruction. Attempts to improve teacher performance led to federal and state efforts to incentivize improved performance through alternative compensation systems. For example, through P.L. 109-149, Congress authorized the Federal Teacher Incentive Fund (TIF) program, which provides grants to support teacher performance pay efforts. In addition, there are various programs at all levels (national, state, and local) aimed at reforming teacher compensation systems. The most recent congressional action in this area came with the passage of the American Recovery and Reinvestment Act of 2009 (ARRA, P.L. 111-5) and, in particular, enactment of the Race to the Top (RTTT) program.

In November 2009, the U.S. Department of Education released a final rule of priorities, requirements, definitions, and selection criteria for the RTTT. The final rule established a definition of an *effective teacher* as one "whose students achieve acceptable rates (e.g., at least one grade level in an academic year) of student growth (as defined in this notice)." That is, to be considered effective, teachers must raise their students' learning to a level at or above what is expected within a typical school year. States, LEAs, and schools must include additional measures to evaluate teachers; however, these evaluations must be based, "in significant part, [on] student growth."

This report addresses issues associated with the evaluation of teacher effectiveness based on student growth in achievement. It focuses specifically on a method of evaluation referred to as *value-added modeling* (VAM). Although there are other methods for assessing teacher effectiveness, in the last decade, VAM has garnered increasing attention in education research and policy due to its promise as a more objective method of evaluation. The first section of this report describes what constitutes a VAM approach and how it estimates the so-called "teacher effect." The second section identifies the components necessary to conduct VAM in education settings. Third, the report discusses current applications of VAM at the state and school district levels and what the research on these applications says about this method of evaluation. The fourth section of the report explains some of the implications these applications have for large-scale implementation of VAM. Finally, the report describes some of the federal policy options that might arise as Congress considers legislative action around these or related issues.

Contents

Contacts

Introduction

Two of the major goals of the Elementary and Secondary Education Act (ESEA), as amended by the No Child Left Behind Act of 2001 (P.L. 107-110; NCLB), are to improve the quality of K-12 teaching and raise the academic achievement of students who fail to meet grade-level proficiency standards. In setting these goals, Congress recognized that reaching the second goal depends greatly on meeting the first; that is, quality teaching is critical to student success. Thus, NCLB established new standards for teacher qualifications and required that all courses in "core academic subjects" be taught by a *highly qualified teacher* by the end of the 2005-2006 school year.[1]

During implementation, the NCLB highly qualified teacher requirement came to be seen as setting minimum qualifications for entry into the profession and was criticized by some for establishing standards so low that nearly every teacher met the requirement.[2] Meanwhile, policy makers have grown increasingly interested in the output of teachers' work; that is, their performance in the classroom and the effectiveness of their instruction. Attempts to improve teacher performance led to federal and state efforts to incentivize improved performance through alternative compensation systems. For example, through P.L. 109-149, Congress authorized the Federal Teacher Incentive Fund (TIF) program, which provides grants to support teacher performance pay efforts. In addition, there are various programs at all levels (national, state, and local) aimed at reforming teacher compensation systems.[3] The most recent congressional action in this area came with the passage of the American Recovery and Reinvestment Act of 2009 (ARRA, P.L. 111-5) and, in particular, enactment of the Race to the Top (RTTT) program.

The ARRA appropriated $4.35 billion to the U.S. Department of Education (ED) for the RTTT program. Since that time, appropriations legislation has continued to fund the RTTT program in FY2011 (approximately $700 million) and FY2012 (approximately $549 million).[4] Eligibility for funds is dependent on four broad areas of school reform outlined by ED:

- adopting standards and assessments that prepare students to succeed in college and the workplace and to compete in the global economy;

- building data systems that measure student growth and success, and inform teachers and principals about how they can improve instruction;

- recruiting, developing, rewarding, and retaining effective teachers and principals, especially where they are needed most; and

- turning around the lowest-achieving schools.

---

[1] According to ESEA Section 9101(11), "The term 'core academic subjects' means English, reading or language arts, mathematics, science, foreign languages, civics and government, economics, arts, history, and geography." For more information on the teacher quality requirements, see CRS Report R42127, *Teacher Quality Issues in the Elementary and Secondary Education Act*, by Jeffrey J. Kuenzi.

[2] According to a study conducted for the Education Department by the RAND Corporation, "By 2006–07, the vast majority [over 90 percent] of teachers met their states' requirements to be considered highly qualified under NCLB." See http://www.ed.gov/rschstat/eval/teaching/nclb-final/report.pdf.

[3] For more information on compensation reform, see CRS Report R40576, *Compensation Reform and the Federal Teacher Incentive Fund*, by Jeffrey J. Kuenzi.

[4] For more information on the funding status of RTTT, see http://www2.ed.gov/programs/racetothetop/funding.html.

Two of the four school reform areas specifically address teacher improvement and teacher effectiveness. By articulating these reform areas, ED has provided an incentive to states to become more systematic about using student data to inform teacher instruction and to measure teacher effectiveness. The latter point is elaborated on in the discussion that follows pertaining to the definition of effectiveness (i.e., "effective teacher") included in ED's RTTT final rule.

In November 2009, ED released a final rule of priorities, requirements, definitions, and selection criteria for the RTTT, which provided details on how states are expected to address the four school reform areas.[5] In the area of teacher effectiveness, the final rule proposed a definition of an *effective teacher* as one "whose students achieve acceptable rates (e.g., at least one grade level in an academic year) of student growth (as defined in this notice)."[6] That is, to be considered effective, teachers must raise their students' learning to a level at or above what is expected within a typical school year. States, LEAs, and schools must also include additional measures to evaluate teachers; however, these evaluations must be based, "in significant part, [on] student growth."

This report addresses issues associated with the evaluation of teacher effectiveness based on student growth in achievement. It focuses specifically on a method of evaluation referred to as *value-added modeling* (VAM). Although there are other methods for assessing teacher effectiveness, in the last decade, VAM has garnered increasing attention in education research and policy due to its promise as a more objective method of evaluation. Considerable interest has arisen pertaining to the feasibility of using VAM on a larger scale—for instance, to meet RTTT program eligibility requirements concerning the evaluation of teacher performance. This report has been prepared in response to numerous requests for information on this topic. While no federal program has specified VAM as the approach that should be used to link teacher performance to student achievement, this examination of the feasibility of implementation and relevant policy implications may generate insights that are helpful in consideration of the use of VAM and alternative approaches to linking student achievement to teacher performance.

The first section of this report describes what constitutes a VAM approach and how it estimates the so-called "teacher effect." The second section identifies the components necessary to conduct VAM in education settings. Third, the report discusses current applications of VAM at the state and school district levels and what the research on these applications says about this method of evaluation. The fourth section of the report explains some of the implications these applications have for large-scale implementation of VAM. Finally, the report describes some of the federal policy options that might arise as Congress considers legislative action around these issues.

---

[5] U.S. Department of Education, "Race to the Top Fund; Final Rule," 74 *Federal Register* 59688-59834, November 18, 2009.

[6] U.S. Department of Education, "Race to the Top Fund; Final Rule," 74 *Federal Register* 59804, November 18, 2009. The definition states, "Effective teacher means a teacher whose students achieve acceptable rates (e.g., at least one grade level in an academic year) of student growth (as defined in this notice). States, LEAs, or schools must include multiple measures, provided that teacher effectiveness is evaluated, in significant part, by student growth (as defined in this notice). Supplemental measures may include, for example, multiple observation-based assessments of teacher performance." Subsequent phases of the RTTT grant competition continued the applicable final requirements and definitions of key terms from the notice of final priorities published November 18, 2009 (see footnote 5).

What is Value-Added Modeling?

VAM is a quasi-experimental[7] method that uses a statistical model to establish a causal link between a variable and an outcome. In education, VAM has been used to establish a link between teachers and the achievement of students within their classroom. This method of modeling is seen as promising because it has the potential to promote education reform and to create a more equitable accountability system that holds teachers and schools accountable for the aspects of student learning that are attributable to effective teaching while not holding teachers and schools accountable for factors outside of their control (e.g., the potential impact of socioeconomic status on student learning).

VAM is actually a flexible set of statistical approaches that can incorporate many different types of models. Some models use student achievement as an outcome and others use student growth. Some models attempt to link teachers to student achievement while other models attempt to link both teachers and schools to student achievement. Although many types of VAM approaches are possible, this report refers to all of these approaches as VAM. There are common elements across these VAM approaches that have policy implications, and these common elements will be explored in the following sections.

VAM is not necessarily equivalent to other "value-added assessment" systems. Some use the term "value-added assessment" to include any method of analyzing student assessments to ascertain growth in learning by comparing students' current level of learning to their own past level of learning.[8] There are some "value-added assessment" systems that do not use VAM,[9] and there are other "value-added assessment" systems that do use VAM.[10] While there are many "value-added assessment" systems, many of them do not use statistical modeling to compare a student's actual growth to a level of expected growth (e.g., one year of academic achievement, average student growth for a school, or some other measure of expected growth). Without comparing actual growth to some pre-defined level of expected growth, a "value-added assessment" system may not be estimating teacher effectiveness. Because the focus of this report is on the estimation of teacher effectiveness—a prominent provision in the RTTT grant competition—only VAM approaches, and not other "value-added assessment" systems, are considered.

The "Teacher Effect"

There are numerous factors that influence student achievement, including past educational experiences, home and neighborhood experiences, socioeconomic status, disability status, the

---

[7] Experimental methods rely on random assignment, such as random assignment of teachers to schools or random assignment of students to teachers. In school settings, random assignment does not occur. Teachers are not hired at random and students are not placed in classrooms at random. For this reason, schools are typically observational settings in which quasi-experimental methods are necessary. A quasi-experimental method uses statistical techniques to approximate experimental conditions; however, this approximation is not perfect, and results will contain a certain amount of uncertainty due to the nonrandom nature of the data.

[8] For example, see http://www.effwa.org/pdfs/Value-Added.pdf.

[9] The Pennsylvania Value-Added Assessment System (PVAAS) measures student growth but does not seem to link student growth to teachers (see http://www.pde.state.pa.us/a_and_t/cwp/view.asp?A=108&Q=108916).

[10] The Tennessee Value-Added Assessment System (TVAAS) links student achievement data and uses VAM to estimate teacher effects (see http://addingvalue.wceruw.org/Related%20Bibliography/Articles/Sanders%20&%20Horn.pdf).

classroom teacher, and so on. VAM recognizes that there are multiple factors that contribute to learning and is therefore designed with the intention of isolating the teacher's effect on student learning. The "teacher effect" is an estimate of the teacher's unique contribution to student achievement as measured by student performance on assessments. It is isolated from other factors that may influence achievement, such as socioeconomic status, disability status, English language learner (ELL) status, and prior achievement. One important feature of the teacher effect is that it is a statistical *estimate* of teacher effectiveness. The teacher effect is simply a statistical value or number, whereas teacher effectiveness is the actual phenomenon being estimated. Another important characteristic of the teacher effect is that it cannot determine *why* a teacher is effective or ineffective, nor does it provide any information on the specific characteristics of what makes a teacher effective. The teacher effect is no more or less than an estimate of the amount of influence a teacher has on the achievement of students in his or her classroom in the content areas being assessed.

Defining a teacher effect is critical to the utility of VAM. If VAM is used to estimate teacher effectiveness, it may be advisable to define the teacher effect consistently across schools, districts, or states, depending on the conclusions one would like to make about teachers (i.e., comparisons of teacher effectiveness across schools, comparisons of teacher effectiveness across districts, or comparisons of teacher effectiveness across states). A teacher effect can be defined in multiple ways depending on two major features: (1) the "plausible alternative," and (2) the other factors in the model (e.g., socioeconomic status, disability status, ELL status, prior achievement, and so on).

The first feature—the "plausible alternative"—defines a teacher effect relative to some other realistic alternative. For example, the teacher effect can be defined relative to the average teacher within a school, average teacher within a district, average teacher within a state, or some other alternative. In current applications of VAM, teacher effects are often estimated relative to the average teacher within a district. Defining the teacher effect in this way may make sense if the goal is to provide information about teacher effectiveness relative to others in the district; however, this definition makes it difficult to make comparisons of teachers across districts within a state. If policy makers pursue the use of VAM approaches, the policy may need to clearly describe the desired comparisons to be made.

The second feature—the other factors in the model—defines how precisely a teacher effect is isolated from other factors that are not attributable to the teacher but can nonetheless affect student achievement. VAM approaches usually include "covariates," which are factors that are thought to affect student achievement but are not attributable to the teacher. For example, one covariate that is often used in VAM is socioeconomic status. By adding covariates in VAM, the model attempts to essentially remove the influence of other factors on student learning. By doing this, the teacher effect is isolated and the modeled teacher effect does not, in theory, reflect student learning that is attributable to these other factors. To maintain consistency in the definition of a teacher effect, VAM approaches may need to use the same covariates across settings.

The use of covariates influences the amount of student achievement that can be directly attributed to a teacher. For example, if a large number of covariates are added to the model, much of a student's achievement may be attributed to these factors, leaving a small amount that can be influenced by the teacher. In this scenario, the teacher effect may be accurately isolated, but the magnitude of the effect may be small. If a small number of covariates are added to the model, much of a student's achievement is available to be explained. In this scenario, the teacher effect

may not be well isolated, but the magnitude of the effect has the potential to be large. If policy makers pursue the use of VAM approaches, the policy may need to clearly describe the covariates of interest that should be included in a model that attempts to isolate the teacher effect.

The use of covariates in VAM is appealing because it allows the teacher effect to more accurately reflect his or her contribution to student performance; however, the use of covariates also introduces several conceptual difficulties for policy. For example, consider the use of socioeconomic status as a covariate. If a student comes from a family of low socioeconomic status, it is likely that this will explain a portion of his or her achievement within the model. Historically, students from families of low socioeconomic status tend to have lower scores on student assessments than students from families of higher socioeconomic status. Should policy assume that socioeconomic status may influence student scores and not make teachers responsible for attaining equitable achievement of students from low socioeconomic status? Or, should policy acknowledge that a factor like socioeconomic status is outside of the control of a classroom teacher and should be taken into consideration when evaluating that teacher? As another example, assume that one of the covariates in the model is disability status. If the model allows a student's disability status to explain a portion of achievement, is that acceptable? Or, should policy expect teachers to be equally effective in teaching students with disabilities and students without disabilities? These are important underlying questions that can inform the use of VAM. Answers to these questions are difficult and depend on the overall goal of education policy.

Components of Conducting a Value-Added Model

Using VAM to estimate teacher effectiveness has the potential to provide clear, useful information to teachers, principals, and policy makers about which teachers are influencing student learning in a positive way. If principals and policy makers can identify effective teachers, they may be able to begin the process of understanding what makes them effective and promote policies and practices that may increase the effectiveness of other teachers. Although the positive potential for using VAM to gauge teacher effectiveness is considerable, VAM is conceptually complex and computationally difficult. The sections below discuss some of these complexities, including the database requirements that must be in place prior to using VAM and the decisions that must be made when calculating a teacher effect. Although there are many statistical issues to consider, the sections below primarily discuss how the statistical complexities of VAM may influence policy decisions regarding the use of VAM to estimate teacher effectiveness.[11]

Database Requirements

To conduct an analysis using VAM, a sophisticated database must be in place, possibly for several years before an analysis can be carried out. The first requirement of a database for VAM is that it must have longitudinal data; that is, the database must include test scores from multiple grades for individual students. Ideally, the test scores would come from the same assessment, and that assessment would have known psychometric properties, such as reliability and validity.[12] Second,

---

[11] For a more comprehensive discussion of statistical issues that influence the estimate of teacher effectiveness using VAM, see Daniel F. McCaffrey, J.R. Lockwood, and Daniel M. Koretz, et al., *Evaluating Value-Added Models for Teacher Accountability* (Santa Monica, CA: RAND Corporation, 2003).

[12] For a discussion of reliability and validity, see CRS Report R40514, *Assessment in Elementary and Secondary Education: A Primer*, by Rebecca R. Skinner.

the database must have variables that link students to teachers. In some cases, this link could be fairly simple. For example, an elementary school teacher who is completely responsible for teaching a class of 20 students could be linked to the assessment scores of these students in a relatively straightforward way. In other cases, this link is not as clear. For example, many students are taught by multiple teachers, such as a regular education classroom teacher and a special education teacher or an English language teacher. In higher grades, students often have multiple teachers—one for each subject. Linking multiple teachers to a student's assessment score is a difficult process that requires some forethought: What fraction of the student's learning should be accounted for by each teacher? In higher grades, which teacher should be responsible for student performance on a reading assessment (e.g., history teacher, English teacher, etc.), given that most students do not explicitly learn "reading skills" in higher grades? Similarly, which teacher is responsible for student performance on a mathematics assessment (e.g., geometry teacher, algebra teacher, trigonometry teacher, etc.), given that a "mathematics" assessment may have items from multiple mathematics courses? Are all teachers responsible? If so, what fraction of student performance should be attributed to each teacher?

A third requirement for databases is general information about the students, teachers, and schools that can be used as covariates in the model. At the student level, information about student race/ethnicity, socioeconomic status, disability status, and ELL status may be included in the database. In addition, any information on the student's family and neighborhood characteristics may be included. At the teacher and school level, information about teacher preparation programs, years of experience, and characteristics of the school may be useful covariates in VAM. In reality, however, information on students, teachers, and schools in large-scale databases is often limited, inaccurate, or missing completely, which may make the use of covariates in VAM inconsistent. Policy regarding the use of VAM may wish to consider which covariates are of interest when estimating teacher effectiveness, and ensure that schools and districts have the capacity to collect this information and report it accurately.

Estimating Teacher Effects

Once an appropriate database is in place, an analyst can construct a specific model using a VAM approach designed to isolate the teacher effect, thus estimating teacher effectiveness. The estimation of a teacher effect requires the analyst to make decisions about the specific model to be used and the covariates to be included. These decisions can affect the results and influence the level of certainty of the teacher effect. The following sections discuss common factors that can influence the calculation of the teacher effect: general issues of statistical modeling; covariates, confounding factors, and missing data; and the use of student assessments.

General Issues of Statistical Modeling

There are many types of VAM approaches that can estimate teacher effectiveness.[13] Models differ along at least two dimensions: (1) how student achievement is conceptualized, and (2) how teacher effectiveness is conceptualized. In terms of how student achievement is conceptualized, some models use a single score on an assessment while others use "growth" or "gain scores" from one year to the next. While there are advantages and disadvantages to both methods, the

---

[13] For example, some common types of VAM approaches include the covariate adjustment model, the gain score model, and multivariate models.

important policy implication to consider is that teacher effects from VAM using a single score and teacher effects from VAM using gain scores may not be directly comparable. Furthermore, the way in which student achievement is conceptualized can affect the magnitude of the teacher effect. In some cases, teachers may be found to be "more effective" using a single score on an assessment than when using gain scores. In other cases, the opposite may be true. Again, an important consideration in the use of VAM is to predetermine the types of comparisons to be made with the results. Teacher effects may not be easily compared across different types of models with different conceptualizations of student achievement.

In terms of how teacher effectiveness is conceptualized, some models consider teachers "fixed effects" while others consider teachers "random effects."[14] Analysts may choose to use either "fixed effects" or "random effects" based on the goal of the VAM analysis. If the outcome of interest is to determine the effectiveness of teachers in a particular school or district relative to each other, it may make sense to consider teachers a "fixed effect." In this scenario, teachers within the same school or district could be compared to each other but not to teachers who were not included in the VAM analysis. If the outcome of interest is to determine the effectiveness of teachers relative to a "hypothetical teacher," it may make sense to consider teachers a "random effect."[15] In this scenario, teachers could be compared more broadly to the hypothetical situation defined by the model. Both methods have advantages and disadvantages in modeling teacher effectiveness. Some researchers have suggested that using a "fixed effect" model may be preferable when using teacher effects within an accountability system;[16] however, some current applications of VAM use a "random effects" model (e.g., the Tennessee Value-Added Assessment System; TVAAS). There are many statistical implications for specifying teachers as either "fixed effects" or "random effects," but, once again, an important policy consideration is the potential to make comparisons of teacher effectiveness. The teacher effect from a "fixed effects" VAM analysis and the teacher effect from a "random effects" VAM analysis may not be easily compared. It may be of interest, therefore, to specify the comparisons of interest before making these modeling decisions.

Covariates, Confounding Factors, and Missing Data

Analysts must also make decisions about the components that constitute the VAM: covariates, confounding factors, and missing data. Decisions about how to include these components can affect the calculation of a teacher effect.

Characteristics of a student or a student's environment that are believed to affect academic achievement but are not attributable to the teacher are called covariates. As discussed above, a covariate is included in a VAM analysis to "factor out" the amount of a student's academic performance for which the teacher is not responsible. By doing so, the teacher effect should be a true representation of the influence of the teacher on achievement and not the influence of so-called "uncontrollable" factors on achievement (i.e., the influence of covariates). Some of the

---

[14] Specifying teachers as "fixed effects" assumes that the observed teachers (i.e., teachers in the current VAM analysis) are the only teachers of interest. Specifying teachers as "random effects" assumes that teachers are sampled from a larger population of interest.

[15] The "hypothetical teacher" would be defined by the analyst for the specific purposes of the model. It could be defined as an average teacher, an effective teacher, or an ideal teacher, depending on the goals of the analysis.

[16] For example, see Daniel F. McCaffrey, J.R. Lockwood, and Daniel M. Koretz, et al., *Evaluating Value-Added Models for Teacher Accountability* (Santa Monica, CA: RAND Corporation, 2003), pp. 64-68.

most relevant covariates in education are factors such as socioeconomic status, disability status, ELL status, and expenditure per student. Although these are commonly discussed covariates, there may be many more covariates that affect student achievement—some of which are not apparent or cannot be easily measured. For example, some research has demonstrated that parental level of education or individual student motivation can influence student achievement, but this information is unlikely to be included as covariates in a VAM analysis because it is generally not available in statewide databases. Furthermore, there may be other covariates that influence student achievement that have not yet been uncovered.

Without knowing all the variables that affect student achievement (and how to measure them), the teacher effect is not completely isolated from any influence of characteristics of a student or a student's environment that is not attributable to the teacher. This introduces bias into the teacher effect due to the influence of unknown factors. That is, a student's learning, or lack thereof, is mistakenly attributed to the teacher when, in reality, the learning may be a function of unmeasured school or community characteristics. Nevertheless, in practical terms, the use of *known* factors (e.g., covariates such as socioeconomic status) to measure teacher effects may be the most accurate method currently available to gauge how much a teacher contributes to student learning. In practice, however, it is possible that even the most accurate method may not be sufficiently precise to provide useful information to teachers and principals due to the unknown factors that are left out of the estimate of teacher effectiveness. This gap between the current state of research and the current needs of practice continues to be negotiated as VAM is used and studied in schools and districts.

Another potential source of bias in the teacher effect may arise due to confounding factors. A confounding factor is something within the culture of the school, community, or neighborhood that can influence the teacher effect. This source of bias may negatively affect teachers who work in low-performing schools where the factors that cannot be measured likely influence student achievement in negative ways. For example, students in low-performing schools may live in communities with more widespread problems that affect student achievement, such as health problems (e.g., malnutrition and undiagnosed vision or hearing problems) or neighborhood factors (e.g., low expectations for academic success or lack of community resources for after-school activities). Although VAM can estimate a teacher effect that reduces the influence of confounding factors, it is difficult to completely isolate the "true" teacher effect from these factors. As such, policy regarding teacher effectiveness may again consider the appropriate comparisons of teacher effects. If teacher effects are to be compared within a school, the influence of confounding factors is less likely to be a problem because most students within a single school will be influenced by similar health and community factors. If teacher effects are to be compared across schools, districts, or states, the influence of confounding factors may introduce bias into the comparisons because of the diversity of health and community factors across schools, districts, and states.

Finally, the issue of missing data can affect the teacher effect. In district-wide or statewide longitudinal databases, there generally is missing data. Due to high levels of student mobility and absence rates, information collected on students may be incomplete. In addition, cultural factors or language barriers may not allow for certain parent and community data to be collected. There are several methods that researchers use to deal with the problem of missing data;[17] however, these methods have not been well tested in the context of VAM.

---

[17] In statistical models, imputation is often used to substitute some value for a missing data point (e.g., hot-deck (continued...)

It is unknown at this time how missing data would affect the teacher effect; however, student data that is missing in a nonrandom way may create bias. If student data is missing on a large number of students who are highly mobile or have numerous absences, this missing data is nonrandom (i.e., students who are frequently absent have a greater chance of having missing data than students who are not frequently absent). Since students who are highly mobile or have numerous absences are likely to perform at a lower level than other students, the missing data may bias the teacher effect depending on how an analyst chooses to deal with missing data. For example, if students who have missing data are excluded from the analysis, the teacher effect may be positively biased and the teacher may appear more effective than his or her true level of effectiveness. Alternatively, if students who have missing data are assigned an "average" value for their missing data, the teacher effect may be negatively biased because the covariates explaining low achievement are not appropriately used in the model.

Use of Student Assessments

Student achievement is measured through the use of assessments.[18] Results of student assessments are used for many purposes, one of which is to evaluate programs and policies. If states choose to incorporate VAM within teacher evaluation systems, it is unclear at this time whether the VAM analyses would be conducted with existing state assessments or whether states would choose to develop new assessments. Currently, states are required by NCLB to conduct assessments in reading, mathematics, and science for grades 3 through 8 and once in high school.[19] If states choose to use existing assessments, VAM can only provide an estimate of teacher effectiveness for teachers who provide instruction in tested subjects (i.e., reading, mathematics, and science) and for teachers of students in the tested grades (i.e., grades 3 through 8 and once in high school). Using existing assessments may exclude a large number of teachers from an evaluation system using VAM (e.g., teachers of students younger than grade 3 or in non-tested secondary grades; teachers of geography, social studies, history, art, music, etc.). In this scenario, teachers within the same school could not all be evaluated using the same system, which may complicate decisions regarding teacher performance, promotion, and tenure. Furthermore, an evaluation system that does not treat all teachers equally has the potential to create internal conflict among a group of teachers within the same school. If states wish to include all teachers in a VAM system, they may need to develop new assessments for currently untested grades and subjects. To create a comprehensive and consistent teacher evaluation system with VAM, states may need to consider the feasibility and cost of developing new assessments in untested grades and subjects.

Regardless of whether states use new or existing assessments, there are several features of assessments in general that may affect their use in a VAM system that estimates teacher effectiveness. One feature of assessments that may complicate the measurement of the teacher effect is scaling. Ideally, scores from different grades in a longitudinal data system would be

---

(...continued)

imputation or regression imputation). Another way some statisticians correct for missing data is to delete all cases that have missing data and exclude them from the analysis.

[18] For more information on assessment in education, see CRS Report R40514, *Assessment in Elementary and Secondary Education: A Primer*, by Rebecca R. Skinner.

[19] For more information on federal testing requirements, see CRS Report RL31407, *Educational Testing: Implementation of ESEA Title I-A Requirements Under the No Child Left Behind Act*, by Rebecca R. Skinner and Erin D. Lomax.

vertically linked to a single scale so that achievement at one grade could be compared to achievement at other grades. In most statewide assessment systems, scores across multiple grades are not vertically linked onto a single scale. If scores are not vertically linked, the calculation of teacher effects across grades may be inconsistent. For example, students may appear to make large gains from 3rd grade to 4th grade, and the teacher effect may be relatively large. The same group of students could appear to make small gains from 4th grade to 5th grade, and the teacher effect would be relatively small. It is possible that the group of students learned the same amount from 3rd to 4th grade as it did from 4th grade to 5th grade; however, the scaling of the test or the items on the test may have been more suited to measuring the gain from 3rd to 4th grade than to measuring the gain from 4th to 5th grade. Thus, without vertical scaling, it is difficult to equate the amount of gain made across grades, and therefore it is difficult to compare the teacher effect across grades.

Another issue related to using student assessment scores in VAM is the timing of the assessment. Currently, state assessments used in accountability systems are administered once per year, typically in the spring. Using this "posttest-only" model of student assessment, a student's gain score would be measured as the difference in achievement in spring of the previous grade to the spring of the current grade. One problem with this model may be the drop in student achievement that occurs over the summer recess. If this drop in achievement affected all students equally, it may not be a problem for VAM. Research has demonstrated, however, that the drop in student achievement during the summer recess may be related to socioeconomic status and ethnicity.[20] In practice, this may translate into negatively biased teacher effects for teachers of minority student groups of low socioeconomic status.

In theory, it may be beneficial to test students twice per year, once in the fall and once in the spring, so that a student's gain score would be measured as the difference in achievement across one grade in school, presumably with one teacher. This "pretest-posttest" model of student assessment may reduce the problem of decreased achievement over the summer recess; however, it introduces more testing into the school year, which may be burdensome. Furthermore, a past evaluation of federal programs found evidence that the "pretest-posttest" model may introduce *more* bias into the teacher effect than the "posttest-only" model.[21] Due to the uncertainty related to "posttest-only" models and "pretest-posttest" models in VAM, it is unclear when school administrators and policy makers should schedule assessments to accommodate VAM.

Another consideration in the use of student assessments to measure teacher effectiveness is the potential for score inflation.[22] Score inflation refers to increases in scores that do not reflect increases in actual student achievement. In the case of score inflation, increases in scores can be attributed to an inappropriate focus on the specific types of items on the test, "teaching to the test," or even cheating. Score inflation is a difficult phenomenon to study, so it is unclear how prevalent score inflation is in educational testing. Increasing the stakes of student achievement, however, may inappropriately incentivize teachers and schools to engage in activities that

---

[20] K.L. Alexander, D.R. Entwisle, and L.S. Olson, "Schools, achievement, and inequality: A seasonal perspective," *Educational Evaluation and Policy Analysis*, vol. 23, no. 2 (2001), pp. 171-191.

[21] Robert Linn, "Assessments and accountability," *Educational Researcher*, vol. 29, no. 2 (2000), pp. 4-14. Linn reported that a number of factors introduced bias into the "pretest-posttest" model. Some of these factors include student selection, scale conversion errors, administration conditions, administration dates compared to norming dates, practice effects, and teaching to the test.

[22] For more information about score inflation, see CRS Report R40514, *Assessment in Elementary and Secondary Education: A Primer*, by Rebecca R. Skinner.

promote score inflation. If estimates of teacher effectiveness are to be used for high-stakes decisions for teachers (such as promotion, compensation, tenure, and dismissal), policy makers may consider implementing certain protections against score inflation (e.g., the use of multiple measures of student assessment, the use of a low-stakes "audit" assessment, etc.).

Practical Applications and Research Results of Value-Added Modeling

Despite the complexities associated with the use of VAM, it is currently used on a limited basis for both teacher and school evaluation. It is not known how many schools or districts have VAM in place; however, the popularity of "value-added" systems continues to grow. Often times, the schools and districts that choose to implement VAM to estimate teacher effectiveness provide limited information on the details of their procedures and their statistical models. There are, however, several large-scale examples of VAM. Two often-cited applications of VAM are the Tennessee Value-Added Assessment System (TVAAS) and the Dallas Value-Added Accountability System (DVAAS). Both TVAAS and DVAAS (pronounced "T-VAS" and "D-VAS") are used as a part of larger, comprehensive evaluation systems that offer monetary incentives for teachers and schools.

Although the available information on the use of VAM is fairly limited, the findings of several research studies may be able to supplement information on VAM and provide policy guidance. The following section discusses VAM in the field, including the current large-scale applications in Tennessee and Dallas. In addition, relevant research findings are reported and discussed in terms of how they may be able to inform future policy surrounding the use of VAM to estimate teacher effectiveness.

VAM in the Field

The TVAAS is perhaps the most widely cited application of VAM. The TVAAS was developed in the mid-1980s by the Tennessee Department of Education and two statisticians from the University of Tennessee.[23] TVAAS is a statewide system that uses student performance on the state assessment to analyze student gain scores.[24] The student gain scores are used to estimate both teacher effects and school effects. The TVAAS system uses prior student records to remove the influence of factors not attributable to teachers (e.g., socioeconomic status or prior achievement); however, the model does not use covariates in the traditional sense.[25] Teachers' records, including the estimate of teachers' effects, are reported only to the necessary school administrators and not to the public. Teachers are typically awarded a salary bonus for high

---

[23] Drs. William L. Sanders and Robert A. McLean.

[24] Tennessee uses the Tennessee Comprehensive Assessment Program (TCAP), which includes both criterion-referenced and norm-referenced items. In the TVAAS system, only norm-referenced items are used to determine gain scores. The gain scores in the TVAAS model are compared to national norms. For more information about criterion-referenced and norm-referenced assessments, see CRS Report R40514, *Assessment in Elementary and Secondary Education: A Primer*, by Rebecca R. Skinner.

[25] The TVAAS uses prior information on each student as a "blocking factor" rather than using individual covariates, such as socioeconomic status, disability status, ELL status, etc. In this model, each student is used as his or her own control. Using a "blocking factor" is another statistical method to factor out the influence of non-teacher variables on student achievement.

performance on the TVAAS. In some cases, principals and teams of teachers are also eligible for monetary awards based on high performance on the TVAAS.

The Dallas Public Schools began developing a ranking system for effective schools in 1984. Over time, the DVAAS was developed by an Accountability Task Force as part of a comprehensive accountability system that incorporated school improvement planning, principal and teacher evaluation, and school and teacher effectiveness. In past years, the DVAAS was used to estimate "Teacher Effectiveness Indices" and "Classroom Effectiveness Indices." The indices represent a composite measurement of multiple outcomes, such as results from qualitative evaluations, student achievement, graduation rates, etc. In its current form, the DVAAS mainly measures "School Effectiveness Indices." The DVAAS uses a VAM that incorporates covariates to control for preexisting differences in student characteristics. The covariates in the DVAAS model include ethnicity, gender, English language proficiency, socioeconomic status, and students' prior achievement. The DVAAS model also controls for school-level variables, such as mobility, crowding, percent minority, and socioeconomic status. Unlike some of the other VAM approaches used in accountability systems, the DVAAS uses multiple indicators, such as student assessment scores, attendance rates, dropout rates, graduation rates, and other indicators selected by the Accountability Task Force. Scores from student assessments, however, are weighted more heavily and contribute more to the overall estimation of school effectiveness than the other indicators.[26] Because the DVAAS primarily measures School Effectiveness Indices, monetary awards are typically awarded for an entire school. The school then decides how to distribute the awards among teachers and staff at the school.[27]

The TVAAS and DVAAS have been in place (in some form) for over 20 years. Although these systems appear to have operated successfully, a perceived lack of transparency has created confusion among accountability analysts and policy makers who have tried to evaluate these systems.[28] It is difficult to determine the exact models that were used to produce the results reported through the TVAAS and DVAAS systems. If policy makers and administrators choose to use these current systems as examples in the use of VAM for teacher effectiveness, more transparency in model specification may be necessary to replicate the results from Tennessee and Dallas. If these systems cannot be replicated reliably, policy makers may not be able to ensure that the estimate of the teacher effect is meaningful, and teachers may not buy in to a system that is perceived to be unreliable. Furthermore, if these systems are not well understood, they may not be able to serve as appropriate models as other districts and states choose to implement VAM programs to estimate teacher or school effectiveness.

---

[26] The DVAAS system uses both criterion-referenced and norm-referenced student assessments.

[27] William J. Webster and Robert L. Mendro, "The Dallas Value-Added Accountability System," in *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?*, ed. J. Millman (Thousand Oaks, CA: Corwin Press, Inc., 1997), pp. 81-99. For additional information about the DVAAS, see http://www.dallasisd.org/eval/research/articles.htm.

[28] For example, see Yeow Meng Thum and Anthony Bryk, "Value-Added Productivity Indicators: The Dallas System," in *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?*, ed. Jason Millman (Thousand Oaks, CA: Corwin Press, Inc., 1997), pp. 100-109; Gary Sykes, "On Trial: The Dallas Value-Added Accountability System," in *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?*, ed. Jason Millman (Thousand Oaks, CA: Corwin Press, Inc., 1997), pp. 110-119; Daniel F. McCaffrey, J.R. Lockwood, and Daniel M. Koretz, et al., *Evaluating Value-Added Models for Teacher Accountability* (Santa Monica, CA: RAND Corporation, 2003), pp. 19-24.

Research Findings

In addition to the use of VAM in states and districts, researchers also have explored the potential use of VAM to estimate teacher effectiveness using data from multiple educational settings. This work may be able to inform the development of policy regarding viable methods for estimating teacher effectiveness because results may have implications for how teacher effects are measured and how teacher effects can be interpreted. In a critical review of the literature on the use of VAM to estimate teacher effectiveness, a team of researchers determined that results generally support the existence of teacher effects; however, the magnitude of teacher effects may have been overstated in some cases. Furthermore, researchers generally expressed concerns about the stability of teacher effects over time.[29]

Researchers who have explored the stability of teacher effectiveness estimates report mixed results. The results suggest that the correlation between the estimate of a teacher's effectiveness from year to year is "modest."[30] Furthermore, the estimated effectiveness of pre-tenure teachers does not necessarily predict their effectiveness post-tenure. For example, one study categorized pre-tenure teachers of reading into quintiles based on their estimated effectiveness; then, the researchers calculated the same teachers' post-tenure effectiveness and categorized the teachers into quintiles. Although many ineffective pre-tenure teachers remained ineffective, 11% of pre-tenure ineffective teachers became effective teachers when measured post-tenure. In the area of mathematics, the estimate of teacher effectiveness seemed to be more stable, with only 2% of ineffective pre-tenure teachers becoming effective post-tenure teachers.[31]

Other researchers have studied the stability of teacher effectiveness estimates and reached similar conclusions. That is, when teachers are ranked by effectiveness and separated into quintiles, the rankings change over time. In general, about one-third to one-fourth of teachers remained within the same effectiveness quintile; however, approximately 10% to 15% of teachers move from the bottom quintile of effectiveness to the top, and an equal number move from the top quintile of effectiveness to the bottom.[32] These results may serve to caution policy makers and school administrators from making tenure and dismissal decisions based solely on teacher effectiveness rankings. It may be possible to use teacher effectiveness rankings as part of an overall evaluation; however, researchers have not studied such evaluation systems.

Although the results suggest that VAM may not accurately rank teachers according to effectiveness, there may be other potential conclusions that can be made using VAM. Some research suggests that VAM can be used to determine whether teacher effectiveness is

---

[29] Daniel F. McCaffrey, J.R. Lockwood, and Daniel M. Koretz, et al., "Literature Review," in *Evaluating Value-Added Models for Teacher Accountability* (Santa Monica, CA: RAND Corporation, 2003), pp. 17-50.

[30] Daniel Aaronson, Lisa Barrow, and William Sander, "Teachers and Student Achievement in the Chicago Public High Schools," *Journal of Labor Economics*, vol. 25, no. 1 (2007), pp. 95-135; Daniel F. McCaffrey, Tim Sass, and J.R. Lockwood, *The Intertemporal Stability of Teacher Effect Estimates*, National Center on Performance Incentives, Working Paper 2008-22, 2008.

[31] Dan Goldhaber and Michael Hansen, *Assessing the Potential of Using Value-Added Estimates of Teacher Job Performance for Making Tenure Decisions*, National Center for Analysis of Longitudinal Research Data in Education Research, Brief 3, November 2008, pp. 1-12.

[32] Daniel F. McCaffrey, Tim Sass, and J.R. Lockwood, *The Intertemporal Stability of Teacher Effect Estimates*, National Center on Performance Incentives, Working Paper 2008-22, 2008; Cory Koedel and Julian R. Betts, *Re-Examining the Role of Teacher Quality in the Educational Production Function*, National Center on Performance Initiatives, Working Paper 2007-03, Nashville, TN, 2007.

significantly different from the average teacher effectiveness. In one study, approximately one-fourth to one-third of teachers could be identified as distinct from the average level of teacher effectiveness.[33] If other studies are able to corroborate these results, this information could have implications for the way policy makers and school administrators use the estimates of teacher effectiveness. If one-fourth to one-third of teachers can be accurately identified as significantly less effective or significantly more effective than the average teacher, policy makers may be able to support some high-stakes decisions for teachers based on VAM in limited contexts.

Researchers who conduct VAM studies generally caution policy makers about making high-stakes decisions based on the measurement of teacher effectiveness. Currently, VAM may not produce estimates that are stable enough to support decisions regarding promotion, compensation, tenure, and dismissal. VAM measures of teacher effects, however, may be useful in a more comprehensive system of evaluation for teachers and schools.

Implications of Large-Scale Implementation

To date, VAM has been used in limited contexts to estimate teacher effectiveness. With the introduction of the RTTT program, however, states may now be incentivized to find new, rigorous methods to evaluate teachers, one of which may be VAM. If states begin to consider the use of VAM to evaluate teachers, there are many questions regarding large-scale implementation that may require some forethought. These questions largely concern the statewide longitudinal data requirements, capacity for data collection and analysis, and transparency of VAM for teacher evaluation.

Data Requirements

There are specific database requirements for VAM analyses. States that pursue the use of VAM may need to have comprehensive statewide longitudinal data systems in place for at least a year (possibly longer) before they can measure teacher effects using student achievement or student growth as an outcome. In addition, if states consider collecting additional student-level information to use as covariates in VAM, there may be new confidentiality and security policies that must be developed and implemented to ensure that students' and teachers' personally identifiable information is protected.

Using VAM to estimate teacher effectiveness may require states to consider the resources, time, and expertise involved with establishing an appropriate database. Although a number of states have already developed statewide longitudinal data systems, either on their own or through an ED grant,[34] it is unclear how many of these data systems currently link teachers to student achievement data. If existing statewide longitudinal data systems do not have this link in place, states may not be able to use data from their current longitudinal data system to estimate teacher effectiveness with VAM. If states choose to create the link between teachers and student achievement from this point forward, it may take a year or more before VAM can be used to

---

[33] Daniel F. McCaffrey, J.R. Lockwood, and Daniel Koretz, et al., "Models for Value-Added Modeling of Teacher Effects," *Journal of Educational and Behavioral Statistics*, vol. 29, no. 1 (Spring 2004), pp. 67-101.

[34] The Institute of Education Sciences (IES), the research arm of ED, administers a grant competition for states to develop statewide longitudinal data systems. For more information, see http://www.nces.ed.gov/Programs/SLDS/.

---

estimate teacher effectiveness.[35] Creating a comprehensive statewide longitudinal data system with teachers linked to student achievement is a large investment; however, the potential for future analyses may extend beyond analyses of teacher effectiveness. States would face a tradeoff between the time and resources necessary to create and maintain the database and the potential information that may be revealed by it.

Capacity

States may also consider whether they have the capacity to conduct VAM analyses in terms of human resources and computing requirements. Measuring a teacher effect with VAM is quite complex computationally and requires an experienced analyst who can make defensible decisions about covariates, confounding factors, and missing data. Although it is possible that accountability analysts may already be trained in this methodology, it is unlikely that most of them possess the necessary skills to conduct VAM in the absence of further training. In addition to human capital requirements, VAM requires sophisticated software to create and run these models.[36] If districts and states choose to use these standard software packages, there is an associated cost with purchasing the software and maintaining licenses for this software. Furthermore, although these software packages are currently available on the market, it is unclear whether they can compute some of the more complex models that are used in research.[37]

Transparency

Due to the complexity of VAM, transparency can be difficult. The estimate of teacher effectiveness using VAM may not be universally accepted if it is not well conceived and communicated to all the appropriate stakeholders. Furthermore, if teacher effectiveness is to be used, in part, for decisions regarding teacher compensation, promotion, tenure, and dismissal, teachers need to understand how their performance will be measured. One way to make the process of estimating teacher effectiveness more transparent is to involve teachers and other school personnel throughout the process. For example, the DVAAS used an Accountability Task Force comprised of parents, teachers, principals, and community and business representatives to design the accountability system for teachers and schools. It may be important for the sustainability of the system to get "buy-in" from teachers and other stakeholders at the beginning of the process. Another way to increase the transparency of VAM may be to allow a second team of analysts to have access to the data in order to corroborate findings. If teacher effectiveness data are to be used for high-stakes decisions, it may be beneficial to have two separate groups of analysts reaching the same conclusions. Replication may increase the scientific rigor of the process and provide additional protection for teachers who are being evaluated using VAM.

The emphasis on transparency of VAM procedures may need to be balanced with an emphasis on student and teacher privacy. As the VAM procedures become more transparent, more information about students and teachers becomes available to analysts or teams of analysts. Although names

---

[35] Sometimes VAM averages the "teacher effect" over several years to make the estimate of teacher effectiveness more reliable. In these cases, it may take three or four years before teacher effectiveness data are reported.

[36] There are several software packages that are available to conduct these analyses. Many researchers currently use hierarchical linear modeling software, which is available from Scientific Software International. In addition, SAS has developed "Schooling Effectiveness—SAS EVAAS K-12" software.

[37] Daniel F. McCaffrey, J.R. Lockwood, and Daniel M. Koretz, et al., *Evaluating Value-Added Models for Teacher Accountability* (Santa Monica, CA: RAND Corporation, 2003), p. 115.

and other personally identifiable information are typically removed from databases before any analysis takes place, states may need to ensure that appropriate privacy policies are in place before they implement VAM. States may also need to consider implementing policies regarding who may have access to data for analysis purposes and who may have access to the results of the data analysis.

Federal Policy Options

Although VAM approaches have been used successfully in district- and state-level contexts to estimate teacher and school effectiveness, research findings related to VAM and implications of large-scale implementation raise issues that may be relevant to the development of federal policy. At this time, it is unclear whether the current applications of VAM can be generalized to a large-scale federal effort or if future research and development is necessary for large-scale implementation. Perhaps other policy alternatives to evaluate teacher effectiveness independent of VAM may be considered (e.g., increasing teachers' and principals' capacity to use student achievement data to inform practice, better use of teacher data to inform teacher evaluation, etc.). If the use of VAM for teacher effectiveness is seen as promising for federal policy, however, there are several short-, mid-, and long-term objectives that may be able to further this goal.

In the short term, federal policy could continue to incentivize states to create databases that can be used for VAM. For example, the RTTT program prohibits eligible states from having any legal, statutory, or regulatory barriers to creating databases that link teachers to student achievement data for the purposes of teacher and principal evaluation. Linking teachers to student achievement data is an essential short-term objective for the use of VAM (or other models of teacher evaluation). Another short-term objective may be to ensure that the student assessments currently in place in elementary and secondary schools are relatively stable and remain in place for a number of years.[38] A consistent measure of student achievement simplifies longitudinal databases and increases the likelihood that VAM can be conducted. In addition to using consistent measures of student achievement, developing consistent measures of potential covariates for VAM analysis may be useful. In some cases, measures of covariates already exist and are collected routinely by schools (e.g., measures of socioeconomic status, disability status, ELL status, etc.). In other cases, however, new measures of covariates of interest may need to be developed (e.g., family characteristic measures, school violence measures, school climate measures, neighborhood measures, etc.), and schools may need to increase the capacity for data collection.

Another short-term objective may be to improve analysts' access to school, district, and state longitudinal databases. In other contexts, analysts have reported difficulty in accessing databases containing high-stakes student achievement data.[39] Although these databases include sensitive information about test scores, analysts who are granted access to actual data may be able to conduct studies on the feasibility of VAM in a typical school context. The federal government may have a role in incentivizing schools, districts, and states to share their longitudinal databases with analysts who are interested in conducting experimental VAM analyses. The potential information gained from granting data access to analysts, however, may need to be weighed against the privacy concerns for students, teachers, principals, districts, and even states. Privacy

---

[38] The current, state-led effort towards common core standards and common assessments may influence states' decisions regarding assessment measures in future years. For more information about the common core standards initiative, see http://www.corestandards.org/.

[39] For example, see Daniel Koretz, *Measuring Up* (Cambridge, MA: Harvard University Press, 2008), pp. 242-245.

policies, confidentiality agreements, and strict protection of identification numbers may be necessary before data access can be granted to analysts outside of the system.

In the mid-term, federal policy could provide startup funding for model demonstration projects of VAM systems in real school contexts. One way to do this may be to scale up current applications of VAM, such as the TVAAS and DVAAS, to other districts or states within the nation. Another way may be to incentivize the development of new teacher accountability systems in which VAM is part of a comprehensive evaluation system. If model demonstration projects of VAM are successful, these models may continue to be scaled up and generalized to new contexts. While the VAM approaches are being generalized, researchers and practitioners may be able to develop "practice guides" that may allow the use of VAM to become more widespread.

Another mid-term objective may be to increase the capacity to carry out VAM in an efficient way. Currently, there is no easily accessible software that can carry out some of the more complicated VAM analyses,[40] and there are few analysts who are qualified to conduct these complicated analyses. The development of more sophisticated, user-friendly modeling software may allow VAM to become more feasible in educational settings. In addition, building human capacity in the use of VAM may be necessary. The federal government has provided funding for capacity building in the past through grants administered by ED. In the current context, grants could be provided for training pre- or post-doctoral fellows in VAM techniques or retraining current accountability specialists in VAM techniques. In addition, the federal government could provide funding to train teachers and principals to make better use of student achievement data and teacher effectiveness data to inform their practice.

In the long term, federal policy may be able to build on successful model demonstration projects of VAM in school settings. In addition, the capacity to conduct this work on a larger scale may be in place. Once VAM is implemented on a larger scale, further evaluation may be warranted. Some researchers advocate using alternative measures of teacher effectiveness to validate the results of VAM.[41] Using alternative measures of teacher effectiveness to validate VAM may potentially lead to more "buy-in" from teachers who are evaluated using VAM. It may also allow teachers, principals, and policy makers to gain a better understanding of what characteristics of teachers make them effective. Currently, a teacher effect can estimate the magnitude of teacher effectiveness; however, the teacher effect cannot, by itself, point to the characteristics of teachers that make them effective. By combining VAM with alternative measures of teacher effectiveness, research and practice may eventually be better able to identify characteristics of effective teachers.

Author Contact Information

Erin D. Lomax
Specialist in Education Policy
elomax@crs.loc.gov, 7-9447

Jeffrey J. Kuenzi
Specialist in Education Policy
jkuenzi@crs.loc.gov, 7-8645

---

[40] Daniel F. McCaffrey, J.R. Lockwood, and Daniel M. Koretz, et al., *Evaluating Value-Added Models for Teacher Accountability* (Santa Monica, CA: RAND Corporation, 2003), p. 115.

[41] See footnote 40.