

11: VALUE-ADDED ASSESSMENT OF TEACHERS

EXECUTIVE SUMMARY

RESEARCH FINDINGS

The Tennessee Value-Added Assessment System (TVAAS) employs a sophisticated statistical methodology to estimate the aggregated yearly growth in student learning, as reflected in changes in test scores in five tested academic subjects. It assumes that changes in test scores from one year to the next accurately reflect student progress in learning. By tracking progress and linking it to schools and teachers, the model asserts that the educational effects of these schools and teachers can be evaluated. Estimates of aggregated gains are used as indicators of how effective teachers and schools have been in raising student performance. Yet, the model's empirical base is weak and fails to document adequately its efficacy as a teacher evaluation instrument. It remains unclear how other variables that may affect achievement as much as teacher effectiveness will determine the evaluation results. Much more research is needed in order to rationally judge the system's strength and weaknesses.

RECOMMENDATIONS

- Develop and implement a program evaluation plan to define and monitor value-added assessment program outcomes. Program evaluation oversight should be maintained by the state and developed and implemented by an independent contractor.
- In order to support and provide guidance for the development and implementation of the program evaluation plan, the state should establish an independent technical panel of experts in measurement, statistics, and educational research methodology.
- The TVAAS database should be made available, along with all technical documentation pertaining to the operations of the TVAAS model, to interested researchers.
- National standards and mechanisms should be developed for the approval of statistical procedures and models to be used in high-stakes accountability systems. Such standards should have the force of a professional code. The task of developing them should be led by the American Educational Research Association (AERA).

11: VALUE-ADDED ASSESSMENT OF TEACHERS: THE EMPIRICAL EVIDENCE

BY HAGGAI KUPERMINTZ
UNIVERSITY OF COLORADO AT BOULDER

The evaluation of teaching has been a major concern in attempts to improve education because “[a] conceptually sound and properly implemented evaluation system for teachers is a vital component of an effective school.”¹ Efforts to develop and implement useful and trustworthy systems of teacher evaluation, however, have frustrated education leaders and policy makers, especially when the evaluation attempted to measure teacher performance by assessing what students have learned. Shrinkfield and Stufflebeam went so far as to declare that “there is no topic on which opinion varies so markedly as that of the validity of basing teacher effectiveness on student learning.”² Various proposals for outcome-based teacher evaluations have been examined under the headings of “process/product” research, school effectiveness research, merit pay and career ladder schemes, public education accountability programs, and private-sector performance contracting.³ Still, persistent substantive and methodological shortcomings of the proposed systems have contributed to “teacher skepticism and growing criticism of attempts to link learning gains to teacher work.”⁴

Recent efforts to reform American education by emphasizing student testing, coupled with significant developments in the statistical modeling and analysis of longitudinal test score data, have sparked a renewed interest in the notion of basing teacher evaluations on measured outcomes of student learning. Whereas traditional school and teacher performance indicator systems have relied on measures of the current level of student achievement, the new systems have shifted their focus to the assessment of year-to-year progress in measured achievement. The

assessment of growth is typically achieved by using some variant of an emerging family of statistical models, collectively known as “value-added assessment.”⁵ The most visible among these contemporary approaches is the Tennessee Value-Added Assessment System (TVAAS), developed in the late 1980s by Dr. William L. Sanders at the University of Tennessee and implemented as the keystone of the Tennessee Education Improvement Act in 1992.

The purpose of this chapter is to describe the TVAAS approach to teacher evaluation and to offer a critical review of the empirical research base that addresses the validity of estimates of teacher effectiveness. It concludes with a set of recommendations intended to strengthen the empirical base of TVAAS and similar programs.

VALUE-ADDED ASSESSMENT RESEARCH

AN OVERVIEW OF TVAAS

TVAAS is the centerpiece of an ambitious educational reform effort implemented by the Tennessee Education Improvement Act of 1992. Inequalities in school funding, followed by a lawsuit brought against the state by a coalition of small rural districts, led to a comprehensive reform of the Tennessee educational system. Under pressure from business, the legislature adopted a strong accountability model that required schools to show concrete evidence of satisfactory year-to-year improvements in student achievement, measured down to the classroom level. Relying on pilot studies that Sanders and his colleagues conducted on the value-added model during the 1980s, the Tennessee legislature embraced the model as its methodology of choice for measuring the performance of students, teachers, schools, and school systems. The legislation defines TVAAS as a “statistical system for educational outcome assessment which uses measures of student learning to enable the estimation of teacher, school, and school district statistical distributions,” and requires that the “system will use available and appropriate data as Value-Added Assessment

input to account for differences in prior student attainment, such that the impact which the teacher, school and school district have on the educational progress of students may be estimated on a student attainment constant basis.”⁶

The TVAAS model, referred to as “the Sanders model’ in some sections of the legislation, employs a sophisticated statistical methodology to estimate the aggregated yearly growth in student learning, as reflected in changes in test scores in five tested academic subjects. Estimates of average student achievement progress are calculated for each school and teacher for each of Tennessee’s school systems. The results are then summarized in a series of reports that show the estimated growth in student achievement attributed to each school system, each school, and each individual teacher in Tennessee. System and school report cards are made public while teacher reports are only shared with their supervisors.

The details of the statistical calculations are too complex to describe in this chapter,⁷ but the idea behind value-added assessment is straightforward. It assumes that changes in test scores from one year to the next accurately reflect student progress in learning. By keeping track of this progress across several years and linking it to the particular schools and teachers who taught the student during that period, the model asserts that the educational effects of these schools and teachers can be evaluated. The larger the aggregated gains attributed to a school or a teacher, the more “value” is said to have been added by them to their students’ learning. Estimates of aggregated gains are used as indicators of how effective teachers and schools have been in raising student performance.

The statistical “mixed model” methodology employed in TVAAS offers several important advantages over competing methods. First, it ensures that all available data will be used in the calculations; other techniques often include in the analysis only students for whom

complete records exist. The statistical calculations take into account the correlation among each student scores across different subjects and across grade levels to provide improved estimates of growth in measured achievement. In addition, the estimation of teacher effects takes into account the amount of data available so that teachers with less data (implying less accurate estimation) are assumed to perform at their system level until more data become available. The model is quite flexible and can be expanded to include different outcome measures and input variables.

Using annual data from the norm-referenced tests that make up the Tennessee Comprehensive Assessment Program (TCAP), schools and school systems are expected to demonstrate progress at the level of the national norm gain (as determined by a national sample of students who took the same tests) in five academic subject areas: math, science, social studies, reading, and language arts. Beginning in 1993, reports have been issued to educators and the public on the effectiveness of every school and school system in Tennessee. Teacher reports are not part of the public record; rather, value-added assessment of teacher effectiveness has been provided only to teachers and their administrators.

TVAAS and Teacher Evaluation

Value-added methodology is increasingly becoming a prominent component in emerging educational accountability systems. The shift in attention from assessing current level of performance to showing systematic progress in learning has enriched and refined the way in which policy makers conceptualize educational outcomes. Consequently, many systems now require schools and teachers to exhibit adequate yearly growth for their students, regardless of how strong or weak the current level of incoming student performance is. TVAAS represents a pioneering effort to implement a comprehensive statewide value-added assessment system to determine the merit of every school system, school, and teacher in fostering student achievement.

Since its inception, TVAAS's advocates have made remarkable claims asserting its effectiveness as an educational accountability tool for teacher evaluation and have promised that TVAAS can provide precise and fair quantitative estimates of the impact of any particular teacher on the academic growth of their students. The system developers have consistently argued that these claims are supported by a strong research base, relying on the massive TVAAS database containing millions of merged longitudinal records of student achievement.

Three major assertions have been offered in support of the TVAAS methodology as a teacher evaluation instrument. In the next section, we will examine the empirical evidence supporting these assertions:

- 1) Teacher effectiveness is by far the most important factor in determining the outcomes of the learning process and TVAAS estimates of teacher effects provide accurate indicators of teacher effectiveness.
- 2) TVAAS estimates of teacher effects measure the independent and unique contribution a particular teacher makes to his or her students' growth, regardless of a student's background.
- 3) TVAAS teacher effects are independent of students' prior ability; therefore teacher effectiveness does not depend on the student's aptitude for learning.

It is clear that any teacher evaluation program possessing these attributes would, indeed, be able to gauge the precise contributions teachers make to students' academic progress – the “value added.” In doing so, such an evaluation program would represent a revolution in accountability. After decades of heroic efforts to disentangle the effects of schooling from the social context in which it is inevitably embedded, the TVAAS system promises to do exactly that. Moreover, it proposes to do so without measuring any of the background variables that have

persistently frustrated generations of researchers and policy makers – variables that led sociologist James Coleman to dismiss the effects of schooling relative to broader social influences more than 30 year ago.⁸ In short, the TVAAS claims that by simply using the student’s past achievement record as a starting point from which to measure progress, and then by keeping track of who teaches the student what, all of the possible influences on this student’s learning – except for those of the teacher, the school, and the school system – can be filtered out or taken into account. No other educational assessment system has ever made such a bold claim. As Education Week’s Jeff Archer noted: “In the current craze for accountability in education, that’s like inventing a state-of-the art mining tool during a gold rush.”⁹ This report now turns to examine the scientific basis that has been offered to validate the above claims.

TVAAS AND PEER-REVIEWED RESEARCH

The scientist’s prime communication tool is the peer-reviewed journal article. Through published articles, innovative methodologies or applications are subjected to rigorous and independent examination by others in the research community. For example, the application of multi-level modeling to the study of school effectiveness (a methodology sharing much in common with TVAAS) has been discussed in countless articles published in educational research journals. Frequently, preliminary findings from the early stages of program development will be presented and discussed in less formal venues. Occasional research reports, working papers, and presentations in workshops or scientific conferences are useful intermediary means to facilitate discussion and to obtain timely feedback. But ultimately, the rigor of the peer-review process is universally accepted in the scientific community (in both the natural and social sciences) as the public forum for the examination of scientific claims. Prominent methodologists Lee Cronbach and Paul Meehl have concluded: “A claim is unsubstantiated unless the evidence for the claim is

public, so that other scientists may review the evidence, criticize the conclusions, and offer alternative explanations. Reference to something ‘observed by the writer in many clinical cases’ is worthless as evidence.”¹⁰ Clearly, the more radical the claim, the more rigorous should be the public examination of evidence, interpretations, and conclusions.

Given the revolutionary nature of the claims advanced by TVAAS developers, it is surprising to find that research findings from TVAAS that specifically pertain to claims regarding teacher effectiveness have been discussed in only three peer-reviewed journal articles, two book chapters, and three unpublished research reports, all of them authored by TVAAS staff. Moreover, out of these, only one journal article and two unpublished reports actually present findings from original empirical studies. Other publications, as well as numerous presentations and newspaper interviews with Sanders and other TVAAS staff, typically repeat these findings and their implications or provide general descriptions of the statistical methodology, program operations, and the variety of reports produced by the system.

The only independent investigations of TVAAS claims and supporting evidence come from two external evaluations of the system. In 1995, teacher concerns about the imminent release of individual teacher reports prompted the Office of Educational Accountability to commission an evaluation study, which indicated several problematic aspects of TVAAS.¹¹ A second external evaluation was initiated in response to the first evaluation and resulted in a report by researchers D. Bock and D. Wolfe¹² dealing with statistical issues, and a companion report by assessment expert T. Fisher addressing implementation and policy issues. The Bock and Wolfe report contains some limited empirical investigations. In addition, two unpublished dissertation studies, one by a former TVAAS staff member and the other by one of the authors of the 1995 evaluation, provide additional analyses.

A note on publications related to the statistical mixed-model methodology employed in the TVAAS is in order here. The theory of mixed-models and related techniques have been among the most productive areas in statistics and have been successfully applied to many practical problems across diverse substantive areas. A vast volume of theoretical expositions and applied research reports documents the validity and utility of mixed-model methodology. The Tennessee Educational Improvement Act makes specific reference to six such publications, furnished by Sanders as support for the TVAAS model.¹³ No one knowledgeable about the issues doubts the soundness of the statistical theory of mixed models, although debate continues about issues such as the efficiency of calculations or estimation algorithms.

The critical point, however, is the validation of any particular application of the general statistical theory. In this regard, reference to the general statistical literature offers no relief. Over the last two decades more and more educational applications have been developed that employ variants of mixed-models methodology, each having to show public evidence for the specific claims, interpretations, and conclusions submitted for consideration. The following section examines the public empirical record concerning TVAAS claims.

TVAAS RESEARCH FINDINGS

This section describes the empirical studies that have been offered to support the three key claims of TVAAS as a system of teacher evaluation, as presented above.

Claim No. 1: *Teachers are by far the most important factor determining the outcomes of the learning process, and TVAAS teacher effects provide accurate estimates of teacher effectiveness.*

One of the most visible TVAAS findings comes from a 1996 study conducted by Sanders and Rivers¹⁴ in two large Tennessee metropolitan school systems. Results were summarized in

an unpublished research progress report. The researchers used longitudinal test scores in mathematics from a cohort of students who started as second graders in 1991-92 and were followed through their fifth grade in 1994-95. Using a simplified version of the full TVAAS model, Sanders and Rivers calculated teacher effects for Grades 3-5 and then arbitrarily grouped teachers into five effectiveness levels according to their relative ranking among their peers. Each group comprised 20% of the teacher sample (referred to by statisticians as “quintiles”). This classification scheme resulted in 125 possible teacher sequences across the three grades – from low-low-low to high-high-high. Fifth grade scores were then used to compare the cumulative effects of seven of these sequences, controlling for second-grade scores. This analysis revealed that the average scores of fifth graders who were assigned the two extreme teacher sequences (low-low-low and high-high-high) differed by about 50 percentile points. Furthermore, for students with comparable teachers in the fifth grade, differences in previous grades’ teacher sequences were still apparent. For example, the differences between the low-low-high and the high-high-high sequences were around 20 percentile points. The researchers concluded that the effects of teachers on student achievement are additive and cumulative, with little evidence for any compensatory effects.”¹⁵

A dissertation study by J. Rivers, one of the authors of the previous study, used a similar analytic strategy. Rivers used fourth-grade math scores and TVAAS teacher effects in Grades 5-8 to predict ninth-grade math competency scores. Although no exact number of students included in the analysis was given, computer outputs presented in the dissertation suggest that the sample size was 2,612. Analyses indicated that teacher effects from Grades 5 to 8 had significant impact on ninth-grade math achievement. In addition a significant interaction effect indicated stronger fifth- and sixth-grade teacher effects on ninth-grade achievement for students

with lower prior achievement. These findings held for the original scale scores on the ninth-grade math competency test and on passing probabilities calculated using a number of different cut scores.¹⁶

In a separate study of a sample of third-, fourth-, and fifth-grade student gains in five subject areas (math, reading, language, social studies, and science) from 1994 to 1995, several context effects were examined in addition to teacher effects: intra-classroom heterogeneity, student achievement level, and class size.¹⁷ Using a simplified version of the TVAAS model, the researchers examined data from 54 school systems in Tennessee. The study addressed classroom context effects. For each grade and subject area, the researchers employed a model that predicted student gains from 12 different factors including school system, classroom heterogeneity, student achievement level (the average of the 1994 and 1995 scores), class size, and various interactions among these terms. The result was a series of 30 separate analyses, one for each grade and subject combination. After comparing the levels of statistical significance of the different effects in the model, the researchers concluded that “the two most important factors impacting student gain are the teacher and the achievement level for the student.”¹⁸ Teacher effects were found to have the largest effect size in two-thirds of the different analyses.¹⁹

Claim No. 2: *TVAAS teacher effects measure the independent and unique contribution a particular teacher makes to his or her students’ growth, regardless of student socioeconomic or ethnic background.*

In an article summarizing research findings from TVAAS, Sanders and Horn²⁰ reported that “the cumulative gains for schools across the entire state have been found to be unrelated to the racial composition of schools, the percentage of students receiving free and reduced-price lunches, or the mean achievement level of the school.” No source or further details were

provided to support this statement. The same assertion has been repeated numerous times in reports and presentations by TVAAS staff, as well as in media coverage of the system. These alleged results have been taken to verify “the contention that by allowing each student to serve as his or her own control (the longitudinal aspect of TVAAS) the inclusion of exogenous co-variables to ensure fairness in the estimates of system, school, and teacher effects is not necessary.”²¹

This contention distinguishes TVAAS from other similar methodologies, as no other contemporary value-added system has reached the same conclusion. Accordingly, such systems typically include an explicit statistical adjustment for competing factors that may influence student progress (over and above their influence on student current level of achievement). In an exposition of value-added indicators of school performance, Robert Meyer explains:

The key idea is to isolate statistically the contribution of schools from other sources of student achievement. This is particularly important in light of the fact that differences in student and family characteristics account for far more of the variation in student achievement than school-related factors. Failure to account for differences across schools in student, family, and community characteristics could result in highly contaminated indicators of school performance.²²

The contention that merely by including in the analysis the student’s previous test scores, the system is able to control adequately for all exogenous influences – without actually measuring them – is a radical departure from the conclusions reached by other researchers, as well as from basic intuitions about schooling. It is counter-intuitive for most educators to assume that student, family, or community resources will have only negligible impact on a student’s rate of progress, even after prior achievement has been accounted for. Extraordinary claims demand extraordinary evidence. Such a radical assertion requires reliable and strong empirical evidence if it is to be trusted to serve as a working assumption for school or teacher evaluations. The only evidence that has been offered to date to support this contention, however, comes from an Value-Added Assessment

unpublished report circulated by the University of Tennessee Value-Added Research and Assessment Center.²³ The document, whose authors are unidentified, displays scatter plots of the percentage of minority students in each of some 1,000 Tennessee schools against the three-year cumulative average gains in each school for the five TCAP-tested subjects, as calculated by TVAAS. The report does not provide any formal statistical analysis of these patterns, leaving the reader to evaluate its conclusions by eyeballing the scatter plots. The report concludes that “the graphs show that the effectiveness of a school cannot be predicted from a knowledge of the racial composition.”²⁴ Yet a closer inspection of the graphs reveal that while they do not display a clear downward trend, schools with more than 90% minority enrollment seem to exhibit lower cumulative average gains. For example, about 70% of the schools with high minority enrollment showed gains that were below the national norm; comparable patterns can be observed for reading, language, and social studies. Similar graphs for school systems reveal an even stronger relationship between average system gains and percentage of students eligible for free or reduced-price lunch, despite the authors’ claim to the contrary.

The Sanders and Rivers 1996 report provides further indirect evidence for the role family background factors may play in influencing student progress. Table 3 of the report gives the frequency with which white and black students, respectively, were assigned to teachers in each effectiveness level. Generally, white students were more often assigned to more effective teachers than were black students. Of white third-grade students in one of the school systems, 15.9% were assigned to teachers in the lowest effectiveness group, compared with 26.7% of the black students who were assigned to similar teachers. In contrast, 22.4% of the white students, and 14.4% of the black students, were assigned to teachers in the highest effectiveness level. These findings echo the well-documented severe inequalities in resources and opportunities

characteristic of the American educational system.²⁵ The link between teacher effectiveness, as measured by TVAAS, and student ethnicity further underscores the fragility of the contention that value-added indicators are unrelated to the racial composition of the student body.

Another dissertation study provides additional demonstration of the relationships between TVAAS value-added scores and exogenous factors. Hu has documented substantial and significant correlations at the school level between per-pupil expenditure, percent minority students, and percent of reduced-price/free lunch students with average TVAAS value-added scores in both math and reading.²⁶ Taken together, these variables explained a sizable proportion of the variability in the value-added three-year average gains.

Claim No. 3: *TVAAS teacher effects are independent of student prior ability; therefore teacher effectiveness does not depend on students' aptitude for learning.*

There is a growing recognition that “[e]ffective instruction begins with what learners bring to the setting.”²⁷ Students bring powerful general and domain-specific ideas, knowledge, and skills to the classroom environment. These initial knowledge and skills are resources for further learning and are ingrained in internal mental representations and dispositions, but also in socially determined patterns of participation, within and outside of school. Prior knowledge and conceptions, both formal and informal, play an important role in student performance and later development.

The study of the relationship between teacher effectiveness and student abilities as determinants of student academic progress faces two major challenges. The first concerns the question of clarifying what role each of these two influences plays in the progress of any individual student. The second concerns the potential confusion that may arise when certain teachers are consistently assigned lower- or higher-ability students. The studies described above

show a rather consistent pattern in which higher-ability students tended to achieve lower gains, regardless of teacher effectiveness level as estimated by TVAAS. This phenomenon has been labeled “a shed pattern,” raising concerns about lack of adequate instructional support for high-ability students. TVAAS presentations also describe different patterns, labeled “reverse shed” (whereby higher gains are made by higher-ability students) and “tee-pee” (whereby higher gains are made by students of average ability). No study to date has examined the relative prominence of these different patterns of growth and initial achievement across classrooms and schools.

Some empirical evidence that illustrates the difficulty of isolating the role of teacher effectiveness from that of student prior achievement comes from data presented in the Sanders and Rivers report. Kupermintz, Shepard, and Linn²⁸ re-analyzed data from Table 1 of the report to demonstrate a sizable correlation between estimates of teacher effectiveness and student prior ability. Among students who hadn’t done well in the past, nearly one-third were then assigned to teachers who were later rated to be the least effective, while among the highest-achieving students, more than one-half were assigned to teachers later found to be “highly effective.” These findings suggest that higher-ability students were assigned to teachers that TVAAS analysis identified as more effective, thereby complicating the claims attributing student progress solely to the teacher.

Similarly, independent analyses conducted by Bock and Wolfe for their evaluation report examined the correlations between students’ average score levels and their average gains in a sample of the Tennessee data.²⁹ Bock and Wolfe have commented:

“Although the magnitude of all of the correlations is less than 0.3, a good number of them are large enough to have implications for the comparison of gains between teachers whose students differ in average achievement level... [A]djustments for expected gain as a function of student score level should be included when the magnitude of the correlation exceeds, say 0.15.”³⁰

When students with higher or lower prior achievement are likely to differ systematically in the amount of progress they can demonstrate relative to their peers due to factors outside of the teacher's control, a potential for biased teacher evaluations exists if some teachers consistently get lower- or higher-achieving students. Under such circumstances, it becomes increasingly difficult to differentiate between learning gains that should be attributable to the teachers and those that reflect the superior aptitude of their students.

A CRITICAL EXAMINATION OF THE EVIDENCE

An examination of the empirical record to date reveals a number of issues in methodology and interpretation that call into question the validity of the major TVAAS claims. The available empirical base consistently documents considerable variations among teachers in estimated TVAAS teacher effects. The studies reviewed in this chapter convincingly demonstrate that students of certain teachers show substantially greater or lower gains on average than the students of other certain teachers. The analyses, however, fail to explain clearly and conclusively why such differences exist.

The causal attribution of student gains to teacher effectiveness, as well as the conclusion that teacher effects are additive, cumulative, and largely irreversible, cannot be dismissed as a plausible hypothesis to explain average achievement gain differentials among teachers. Other untested hypotheses remain equally plausible, however. If observed patterns of academic progress are a function of complex interactions between instructional practices, student readiness to learn, and school and community context factors, then teacher effectiveness should be seen as one component of estimated teacher effects. Because there have been no studies that credibly isolate teacher effects from these other factors, however, the question remains open. The current TVAAS model, by default, attributes to the teacher *all* the effects of the possible factors, which

in reality are probably confounded with or interacting with teaching practices.

At least two studies have documented the importance of student ability on academic progress and suggested more complex interactions with teacher effects than has been acknowledged by the authors. The issue of the relationships between student aptitude, student actual achievement, and academic progress is a complicated one. “Shed patterns,” for example, indicate that students with higher prior achievement tend to exhibit smaller gains. Such patterns may result from inefficient instruction for these students, but they may also reflect statistical artifacts like “ceiling effects” (high ability students scoring at the highest levels of the measurement scale, leaving little room for observed improvement) or “regression to the mean” (the statistical tendency of extreme scores to converge to the mean in subsequent measurements). It is unclear to what extent such artifacts may affect the TVAAS model. A contrasting growth pattern will be expected if students lacking in aptitude also tend to make lower gains. Because of the high correlation between aptitude and achievement these trends would result in contradictory findings. Furthermore, potential systematic inequalities in the assignment of teachers to students would complicate the problem further if teachers are systematically assigned to students with different potential to show progress, regardless of a teacher’s efforts. To date, no systematic analysis addresses this crucially important question. Indeed, TVAAS has shown inconsistent results in this regard. It is not sufficient to insist that in any case the pattern of gains exhibited by students differing in ability reflects the efficacy of teachers in addressing low- or high-achieving students. Rival hypotheses remain competent contenders. An informative validation would document teacher practices to determine how well they address students with different abilities, and the extent to which TVAAS estimates reflect these practices.

A major source of confusion appears to be the circular nature of the line of argumentation

that attempts to define teacher effectiveness in terms of estimated teacher effects. This has been noted by other researchers.³¹ Statements like “differences in teacher effectiveness were found to be the dominant factor affecting student academic gain”³² are highly misleading. It leaves the reader with the impression that “teacher effectiveness” and “student academic gain” are two different variables and that the former predicts the latter. In fact, teacher effectiveness is *defined* by student academic gain. The only defensible interpretation of the various findings is that teachers vary as to the extent of their students’ average academic gain. Causal attribution, almost by default, of this variability to “teacher effectiveness” has to remain suspect until further validation studies become available. At a minimum, such studies should employ independent measures of teacher effectiveness, such as teaching practices, supervisor evaluations, scores from teacher tests, and so on.

SUMMARY AND RECOMMENDATIONS

The idea of evaluating schools and teachers on the basis of the “value added” to students’ education each year has wide appeal for policy makers. Instead of ranking schools or teachers from best to worst, the intention is to monitor the amount of gain in student achievement from one grade to the next. This approach has obvious advantages over the traditional alternatives when coupled with a sophisticated statistical modeling apparatus capable of handling massive cumulative longitudinal data. Technical and methodological sophistication, however, are only part of the full array of considerations that form a comprehensive evaluative judgment.

Ultimately, the value of any proposed methodology and the information it produces heavily depend on the soundness of the claims made by the system’s advocates. A validity argument assembles and organizes the empirical evidence as well as the logical line of reasoning linking the evidence to favored inferences and conclusions. A useful and valid model must begin with a

sound theory.

Learning and development are arguably the most complex and intriguing phenomena explored by social science. An emerging learning science has started to make tentative inroads into understanding the many facets of the interactions of teaching and learning. Providing effective teaching to support and cultivate learning is therefore the most complex design problem facing educators. Yet the TVAAS model represents an overly simplistic description of teaching and learning. It is in stark contrast to a very rich body of research on learning and teaching that has demonstrated the enormous importance of student learning histories and contextual factors on the rate of academic progress. It seems to ascribe to the teacher an unrealistic responsibility for student learning. No doubt, teachers can make a critical difference in student academic growth, but so can student preparation, the support they receive outside of school (tutoring and summer school are obvious examples), the school context, and the community context – that is, the resources available to the school. Not measuring these factors does not mean they don't have important effects, only that their effects don't get a chance to show through. Unmeasured factors could potentially bias the evaluation results to the extent they play a role in determining learning outcomes. Teachers who operate in a supportive environment at the school and community levels, where students have access to a wealth of resources and enriched learning experiences, will likely be evaluated more favorably than their similarly able counterparts who struggle with harsher conditions. The TVAAS model controls only for prior student achievement, yet empirical evidence is lacking to document the assertion that prior achievement may serve as a reasonable proxy for all the other factors that matter to student learning.

The simplicity of the TVAAS model poses an interesting policy paradox. An implicit assumption of the model is that teachers, not students, are responsible for learning and that

teachers hold the responsibility to produce measurable progress in learning outcomes. This is a common theme in interpreting TVAAS results. This assumption contradicts an opposite emphasis on student accountability. If indeed, as TVAAS has purported to show, “teachers are the single most important factor affecting student growth,” then a student’s failure to pass a gateway or graduation exam is mainly the responsibility of the teacher. This passive view of students seems unrealistic and may send conflicting messages to teachers and students.

An examination of the TVAAS model’s empirical base shows that much more research is needed in order to arrive at a rational judgment of the system’s strength and weaknesses. Currently, only a few sketchy empirical studies have been relied upon to substantiate strong claims of the system’s merits. In light of this weakness, the recommendations below are intended to establish proper mechanisms to ensure the validity and usefulness of current and future educational accountability systems that use the TVAAS model.

- Develop and implement a program evaluation plan to define and monitor value-added assessment program outcomes. The plan should specify the intended goals for the program and how they will be measured. Periodic program evaluation reports should be required to monitor program performance. The plan should also include specifications of potential unintended consequences and a mechanism to ensure that they are kept at an acceptable minimum. Program evaluation oversight should be maintained by the state and developed and implemented by an independent contractor.
- In order to support and provide guidance for the development and implementation of the program evaluation plan, the state should establish a technical panel of experts in measurement, statistics, and educational research methodology. The panel would be

asked to provide routine input into the evaluation process and help policy makers with the technical issues. The panel should also be actively involved in the design and analysis of the various studies and data analyses performed by the independent contractor.

- The TVAAS database should be made available, along with all technical documentation pertaining to the operations of the TVAAS model, to interested researchers. The state should seek proposals from independent researchers for studies that address the validity of the major claims advanced by TVAAS developers. The technical panel can provide input as to the merit of the various proposals and suggest improvements.
- National standards and mechanisms should be developed for the approval of statistical procedures and models to be used in high-stakes accountability systems. Such standards should have the force of a professional code. The task of developing them should be led by the American Educational Research Association (AERA).

REFERENCES

- 1 J. H. Stronge, "Improving Schools Through Teacher Evaluation," in *Evaluating Teaching: A Guide to Current Thinking and Best Practice*, ed. J. H. Stronge (Thousand Oaks, CA: Corwin Press, 1997), 1.
- 2 A. J. Shinkfield and D. L. Stufflebeam, *Teacher Evaluation: Guide to Effective Practice* (Boston: Kluwer Academic Publishers, 1995).
- 3 See J. Millman and H. D. Schalock "Beginnings and Introduction," in *Grading Teachers, Grading Schools: Is Student Achievement A Valid Measure?* ed. J. Millman (Thousand Oaks, CA: Corwin Press, 1997), 6-7. Table 1.1 for a summary
- 4 Ibid., 7.
- 5 For reviews, see H. Meyer, "Value-Added Indicators of School Performance," in *Improving America's Schools: The Role of Incentives*, eds. E. A. Hanushek and D. W. Jorgenson (Washington, DC: National Academy Press, 1996), 200.
L. Saunders, "A Brief History of Educational 'Value Added': How Did We Get to Where We Are?" *School Effectiveness and School Improvement* 10, no. 2 (1999): 233-256.
- 6 State of Tennessee, Education Improvement Act (1992), §49-1-603
- 7 For a detailed description of the TVAAS model, see W. L. Sanders, A. M. Saxton and S. P. Horn, "The Tennessee Value-Added System: A Quantitative Outcomes-Based Approach To Educational Assessment," in *Grading Teachers, Grading Schools: Is Student Achievement A Valid Measure?* ed. J. Millman (Thousand Oaks, CA: Corwin Press, 1997), 137-162.
- 8 J. S. Coleman et al., *Equality of Educational Opportunity* (Washington, DC: U.S. Government Printing Office, 1966).
- 9 J. Archer, "Sanders 101," *Education Week* 18, no. 34 (5 May 1999): 26-28.
- 10 L. J. Cronbach and P. E. Meehl, "Construct Validity In Psychological Tests," *Psychological Bulletin* 52 (1955): 281-302.
- 11 A. P. Baker and D. Xu, *The measure of education: A review of the Tennessee Value Added Assessment System* (Nashville: Tennessee Comptroller of the Treasury, Office of Education Accountability, 1995).
- 12 R. D. Bock and R. Wolfe, *Audit and Review of the Tennessee Value-Added Assessment System (TVAAS): Final Report, 1996* (Nashville: Tennessee Comptroller of the Treasury, Office of Education Accountability, 1996).
- 13 State of Tennessee, Education Improvement Act (1992), §49-1-604. Mixed model methodologies.
- 14 W. L. Sanders and J. C. Rivers, *Cumulative And Residual Effects Of Teachers On Future Student Academic Achievement, Research Progress Report* (Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center, 1996).
- 15 Ibid.
- 16 It is interesting to note that an error message from the computer run presented in Table 2 of Rivers' dissertation suggests that estimates from the analyses may be biased. A possible reason for this error is the inclusion of highly correlated variables in the same analyses.
- 17 S. P. Wright, S. P. Horn, and W. L. Sanders, "Teacher And Classroom Context Effects On Student Achievement: Implications For Teacher Evaluation," *Journal of Personnel Evaluation in Education* 1, no. 1 (1997): 57-67.
- 18 Ibid.
- 19 This study used a statistical technique to compare the magnitude of the different factors affecting student achievement, citing Rosenthal as a supporting reference. However, Rosenthal suggests the technique for comparing results *across different studies* (a methodology known as "meta-analysis) and not for comparing Value-Added Assessment

-
- the magnitude of effects within the same study. The authors offer no justification for this unconventional application of the technique.
- 20 W. L. Sanders and S. P. Horn, "Research Findings From The Tennessee Value Added Assessment System (TVAAS) Database: Implications For Educational Evaluation And Research," *Journal of Personnel Evaluation in Education* 12, no. 3 (1998): 247-256.
 - 21 W. L. Sanders, "Value-Added Assessment," *The School Administrator* 55, No. 11 (December 1998).
 - 22 H. Meyer, "Value-Added Indicators of School Performance," in *Improving America's Schools: The Role of Incentives*, eds. E. A. Hanushek and D. W. Jorgenson (Washington, DC: National Academy Press, 1996), 200.
 - 23 *Graphical Summary of Educational Findings From the Tennessee Value-Added Assessment System* (Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center, 1997).
 - 24 Ibid.
 - 25 L. Darling-Hammond and L. Post, "Inequality in Teaching and Schooling: Supporting High-Quality Teaching and Leadership in Low-Income Schools," in *A Notion at Risk: Preserving Public Education as an Engine for Social Mobility*, ed. R.D. Kahlenberg (The Century Foundation / Twentieth Century Fund Inc. 2000).
 - 26 D. Hu, "The Relationship of School Spending and Student Academic Achievement When Achievement is Measured by Value-Added Scores" (Ph.D. diss., Nashville, TN: Vanderbilt University, 2000).
 - 27 *How People Learn: Brain, Mind, Experience, and School*, eds. J. D. Bransford, A. L. Brown and R. R. Cocking, (Washington, DC: National Academy Press, 1999), xvi.
 - 28 H. Kupermintz, L. Shepard, and R. Linn, "Teacher Effects as a Measure of Teacher Effectiveness: Construct Validity Considerations in TVAAS (Tennessee Value Added Assessment System)," in *New Work on the Evaluation of High-Stakes Testing Programs*, chair D. Koretz, Symposium at the National Council on Measurement in Education (NCME) Annual Meeting, Seattle, WA, April 2000.
 - 29 Bock and Wolfe.
 - 30 Ibid., 27
 - 31 G. W. Bracey, "The Ninth Bracey Report On the Condition Of Public Education", *Phi Delta Kappan* 81,no. 2 (October 1999): 147.
L. Darling-Hammond, "Toward What End? The Evaluation of Student Learning for the Improvement of Teaching," in *Grading Teachers, Grading Schools: Is Student Achievement A Valid Measure?* ed. J. Millman (Thousand Oaks, CA: Corwin Press, 1997).
 - 32 Sanders and Horn, 251.