

Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research

Abstract

The Tennessee Value-Added Assessment System determines the effectiveness of school systems, schools, and teachers based on student academic growth over time. An integral part of TVAAS is a massive, longitudinally merged database linking students and student outcomes to the schools and systems in which they are enrolled and to the teachers to whom they are assigned as they transition from grade to grade. Research conducted utilizing data from the TVAAS database has shown that race, socioeconomic level, class size, and classroom heterogeneity are poor predictors of student academic growth. Rather, the effectiveness of the teacher is the major determinant of student academic progress. Teacher effects on student achievement have been found to be both additive and cumulative with little evidence that subsequent effective teachers can offset the effects of ineffective ones. For these reasons, a component linking teacher effectiveness to student outcomes is a necessary part of any effective educational evaluation system.

Introduction

In the mid-1980s, the level of concern for the state of education in the United States rose across the nation in the wake of the publication of *A Nation at Risk*. In 1989, President George Bush convened an Education Summit with the governors of the fifty states to consider the problems facing education in America and to develop a plan of action for addressing these problems. To this end, the governors, with input from the White House and from experts in the field of education, developed six National Education Goals. How these goals were to be accomplished was left up to the individual states. In those states in which such legislation had not already been adopted, governors returning from the Education Summit began challenging their legislators to develop legislation that would enable the achievement of the National Goals for Education. The legislation enacted as a result of this impetus varied radically from state to state, but despite the differences, reform legislation held in common the call for higher academic standards and greater accountability linked to assessment of educational outcomes.

In Tennessee, a major attempt to improve educational opportunities for students predated the 1989 educational summit. The Comprehensive Education Reform Act (CERA) was enacted under Governor Lamar Alexander in 1984. CERA included a major increase in educational spending and a Career Ladder system for teachers. The teacher evaluation system developed to assess candidates for the upper levels of the Career Ladder was performance based and included an extensive portfolio in addition to three intensive days of on-site observation and dialogue with state-trained evaluators. The portfolio was dropped after the first year due to extremely negative teacher reaction and was replaced by an enhanced dialogue. Many teachers believed that the Career Ladder was bogus because the observations and dialogues could be "performances" rather than illustrations of actual teaching expertise, even though two of the three observations were supposedly unannounced. Eventually, in 1997, the Career Ladder was frozen. Teachers who had already achieved Career Ladder status would retain it, but no new evaluations would take place.

Discussions were initiated in 1989 that would lead Tennessee to its second major education legislative initiative in less than a decade. A law suit filed by a group of the smaller school districts in the state (was the impetus for the second legislative initiative.) It contended that it was the state's responsibility under the state constitution to provide equal funding across districts to ensure an equal educational opportunity for all students. On March 11, 1992, after two years of discussion and debate, the Education Improvement Act (EIA) was signed into law by Governor Ned McWherter. This legislation included another major increase in funding for education in the state, which, in turn, required a second increase in the state's sales tax in less than ten years. Legislators in both parties demanded a strong accountability provision be included in the act to ensure that the new monies would be spent to improve student academic achievement. The Tennessee Value-Added Assessment System (TVAAS), along with

measures including promotion, attendance, and dropout rates of individual schools, would provide information to form the base for the state's new educational accountability system.

TVAAS, referred to in the EIA as the *Sanders model*, was the methodology designated to ascertain the effectiveness of school systems, schools, and teachers in producing academic growth in Tennessee students, thereby linking student academic outcomes to educational evaluation for the first time. TVAAS required the convergence of a statewide testing program, which tests each student each year in several academic subjects, and an unprecedented application of a statistical approach that enables a massive multivariate longitudinal analysis even with fractured student records, which are always present in realworld student achievement data.

Initially, much skepticism was expressed concerning both the use of student achievement data in this manner and the application of statistical mixed-model theory and methodology to yield fair, objective, and unbiased estimates of the system, school, and teacher effects on the rate of academic progress of populations of students. However, detailed external reviews from both the statistical and educational evaluation communities have confirmed that the properties of the TVAAS results are as claimed by Sanders and his associates.

Beginning in 1993, reports have been issued to the educators and public of the State of Tennessee on the effectiveness of every school system and school that serves elementary students in any of the third through eighth grades and for selected high school math courses. Even though teacher reports are not a public record, TVAAS has provided teachers and their administrators with estimates of teacher effectiveness, as well. The EIA states that TVAAS data will be a part of the evaluation of those teachers for whom such data are available. However, how large a part it will be is left to the discretion of the evaluative body. It cannot be the only source of data in a teacher's evaluation.

Recently, a new State Framework for Local Evaluation of teachers has been adopted. Under the new Framework, teachers and their principals devise a professional development plan that must correlate with the plans developed for their schools, just as the school plans must correlate with the plans for the system. The connection between professional growth of the teacher and school improvement must be explicit. Thus, student academic growth, an integral part of virtually every school improvement plan, becomes a part of successful teacher evaluation for the first time on a statewide basis. TVAAS teacher reports are providing the data from which individual professional development plans are drawn. Successful teacher evaluations include an assessment as to how well the teacher has accomplished the goals and professional growth incorporated in the professional development plan. Subsequent TVAAS teacher reports may also furnish evidence that improvement has or has not been achieved.

Beyond the use of the TVAAS database for its assessment and evaluation purposes, it has become a launching pad for research to investigate both accelerators and inhibitors of academic growth of student populations. Important educational questions are now being addressed with a research efficiency that only millions of longitudinal merged records allow. This article is an attempt to summarize the most important findings to date.

A Brief Description of TVAAS

The Tennessee Value-Added Assessment System is a statistical method of determining the effectiveness of school systems, schools, and teachers. TVAAS uses statistical mixedmodel theory and methodology to enable a multivariate, longitudinal analysis of student achievement data. (For a detailed description of TVAAS methodology, see Sanders, Saxton & Horn, 1997.) These data include student scores on (1) the Tennessee Comprehensive Assessment Program (TCAP), a group of tests in five subject areas (math, science, social studies, reading, and language arts) administered annually to all Tennessee students in grades three through eight, and (2) end-of-course tests in high school subjects, currently in five mathematics subjects. The statistical models used in TVAAS are not restrictive as to the indicator variables that can be employed in the process. Rather, any variables linear in their metrics, highly correlated with curricular objectives, and possessing appropriate measurement sensitivities could be used.

Each student's test data are accumulated over time and are linked to that student's teacher(s), school(s), and school system(s). TVAAS utilizes the scaled scores students make over time to model their learning patterns. By taking advantage of the longitudinal aspect of the data, it is possible to note when the normal pace of academic growth deviates. By following growth over time, the child serves as his or her own "control." This enables the partitioning of school system, school, and teacher effects free of the exogenous factors that influence academic achievement and that are consistently present with each child over time.

The massive database comprising these data, currently in excess of 5 million records, makes it possible to utilize many years of information in determining the effectiveness of educational institutions and personnel. In addition to the robustness of the statistical process, further provisions are made within the EIA to ensure fairness. These provisions include basing assessment on at least three years of data and requiring that schools, school systems, and teachers cannot be assessed solely on the basis of TVAAS.

The primary purpose TVAAS serves in the EIA is to provide information for summative evaluation regarding how effective a school, system, or teacher has been in leading students to achieve normal academic gain over a three-year period. TVAAS reports, issued annually, include information on student gains for each subject and grade for the three most recent years as well as the three-year average gains. The cumulative average gain is the primary indicator by which success is measured. According to the EIA, the standards to which school districts and schools within each district are held accountable are expressed in terms of academic gains instead of an expectation set in terms of absolute scores as has been the case in other states. The increasingly important emphasis on student outcomes in the assessment of educational effectiveness is clear from this legislative provision.

If the purpose of educational evaluation is to improve the educational process, and if such improvement is characterized by improved academic growth of students, then the inclusion of measures of the effectiveness of schools, schools systems, and teachers in facilitating such growth is essential if the purpose is to be realized. Of these three, determining the effectiveness of individual teachers holds the most promise because, again and again, findings from TVAAS research show teacher effectiveness to be the most important factor in the academic growth of students.

Even though the driving force for the creation of TVAAS was for summative evaluation, the real power of the process lies in its ability to serve as a data source for formative evaluation and for educational research. Schools, systems, and teachers receive reports detailing their effectiveness with students of different achievement levels so that they may more effectively plan their curricula, pedagogy, and special programs. This information has been found to be invaluable by many teachers and school administrators involved in curricular planning, program evaluation, and developing strategies to meet the needs of students with differing academic attributes and abilities. The reports allow school systems to pinpoint grade and subject problems and successes and to direct efforts and resources accordingly. School reports inform principals not only about how effective the fourth-grade math program is, for example, in regard to enhancing student academic gain but also whether it is equally effective in encouraging such growth in its high achievers as well as in its low-achieving students. Similar information is provided to each classroom teacher along with the formal teacher reports, which have been issued annually since 1996.

In addition to the analysis and reporting of the effects of educational entities for both summative and formative purposes, research initiatives are a major priority. Already many significant findings have resulted from analysis of TVAAS data. Research to date has focused on factors impacting the academic growth of students.

Research Findings

Race, Socioeconomic Factors, and Academic Gain

For grades three through eight, 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. These consistent findings have verified the contention that by allowing each student to serve as his or her own control (the longitudinal component of TVAAS), the inclusion of exogenous covariables to ensure fairness in the estimates of system, school and teacher effects is not necessary. Schools, systems, and teachers who do best under TVAAS are those who provide academic growth opportunities for students of all levels of prior academic attainment.

The Building Change Effect (excerpted from Sanders, Saxton, Schneider, Dearden, Wright, Paul & Horn, 1994)

When the first TVAAS reports of school system effects on student academic growth were being developed in late 1992, it was observed that certain systems had a noticeable drop in gain for all subjects at certain grades. However, the point of retarded gains varied from system to system. After further examination at the system level, it was hypothesized that the entry point into the receiving school could be causing the retardation in growth.

Subsequently, an analysis to test this hypothesis using the records of all Tennessee students was initiated. Since students' records are matched and merged over all systems in the state, school change patterns can be ascertained. School configurations across the state vary enormously, representing nearly all possible combinations of grades two through eight.

The population of Tennessee students relative to transfer status were presented as three groups—those who attended the same school in two contiguous years, those who changed schools and entered the new school at the lowest grade level offered in the receiving school, and those who changed schools but entered the new school at any grade other than the lowest offered. The study found that the negative effect of student transfer to any grade other than the lowest grade offered in the new school is negligible for transfers prior to the seventh grade. However, the effects of transfer are entirely different when students transfer to the lowest grade in the new school building. When populations of students change buildings, regardless of the grade level, the loss in expected academic gain is dramatic. The mean gain for students who transfer to the lowest grade of their new school was measurably lower in all twenty-five subject-grade combinations. Severe retardation in gains was most pronounced in grades six and seven, the grades at which many school systems routinely transfer students en masse to middle school or junior high.

Factors Impacting Academic Gain Among High-Achieving Students

TVAAS data have shown that some schools and, indeed, some school systems have been successful in addressing the needs of students at all levels of achievement as evidenced by their ability to consistently show normal—and sometimes, exceptional—academic progress for students of all academic abilities. However, there is evidence that, aggregated statewide, students at the highest levels of achievement show somewhat less academic growth from year to year than their lower-achieving peers. In a 1997 study, Wright, Horn, and Sanders (1997, p. 63) investigated simultaneously the effects of teachers, intraclassroom homogeneity, and class size on achievement gain. The analyses revealed that

the two most important factors impacting student gain are differences in classroom teacher effectiveness and the prior achievement level of the student. The teacher effect is highly significant in every analysis and has a larger effect size than any other factor in twenty of the thirty analyses. A notably nonsignificant factor was class size. . . . The main effect for heterogeneity was statistically significant in only two of the thirty analyses, approximately the number that would be expected to occur by chance.

Regarding the factor of prior achievement level of the student, Wright et al. (1997, pp. 65-66) note that

no universally applicable pattern emerges, but it is worth noting that out of the twenty-six analyses in which achievement level was significant, the largest gains occurred in the lowest achievement group twelve times, in one of the two middle groups eight times, and in the highest group six times. Similarly, the smallest gains occurred in the highest achievement group fifteen times, in one of the two middle groups six times, and in the lowest group five times. In other words, there is a disturbingly common but not universal pattern for the best students to make the lowest gains. . . . Disproportionately, high-scoring students were found to make somewhat lower gains than average and lower-scoring students.

In a discussion of their findings, Wright et al. (1997) consider explanations for this shed pattern," so called because academic gains drop off as achievement level rises, creating a downward slope like a shed roof. They conclude (p. 66) that

possible explanations include lack of opportunity for high-scoring students to proceed at their own pace, lack of challenging materials, lack of accelerated course offerings, and concentration of instruction on the average or below-average student. This finding indicates that it cannot be assumed that higher-achieving students will "make it on their own."

While this study discovered that "student academic level was found to be significantly related to academic progress" (p. 66), one factor was shown to be far more significant in predicting student academic growth. The "shed" patterns, while predictable to some degree by the previous academic standing of the student, were much more significantly related to the effectiveness of the teacher (Wright et al., 1997, p. 66).

Differences in teacher effectiveness were found to be the dominant factor affecting student academic gain. The importance of the effects of certain classroom contextual variables (e.g., class size, classroom heterogeneity) appears to be rather minor and should not be viewed as inhibiting to the appropriate use of student outcome data in teacher assessment. These results indicate that any realistic teacher evaluation process should include as a major component a reliable, valid measure of a teacher's effect on student academic growth. If the ultimate goal is the improvement in academic growth of student populations, one must conclude that improvement of student learning must begin with the improvement of relatively ineffective teachers regardless of the student placement strategies deployed within a school.

Finally, it is worth noting that results from the ACT and annual writing assessment data from all students in grades four, eight, and eleven are also linked to the TVAAS database. In another TVAAS study, which considered only the top quartile of Tennessee eighth graders, huge differences in mean ACT scores, obtained from these students four years later, were observed among school districts, indicating the enormous difference in effectiveness that exists among districts beyond elementary school (*Graphical Summary*, 1995).

Residual Effects of Teachers on the Academic Gains of Students (summarized from Sanders & Rivers, 1996)

In a 1996 study entitled *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement*, Sanders and Rivers examined the question, "does the influence of a teacher's effectiveness in facilitating academic growth for his/her students continue when these students advance to future grades?" (Sanders & Rivers, 1996, p. 1). The question was explored using data from two large Tennessee metropolitan school systems over a period of four years. For this study, only data pertaining to mathematics were considered.

Sanders and Rivers estimated teacher effects for teachers who taught mathematics in grades three, four, and five. Subsequently, for the purpose of the study, the teachers' effects were divided into five quintiles, with the least effective teachers comprising the first quintile and the most effective teachers the fifth. Student records were linked to those of their teachers, rendering it possible to trace the progress of the students through sequences of teachers identified by their effectiveness.

Sanders and Rivers found that "by looking at sequences in which the fifth-grade teachers were comparable in terms of effectiveness, it is possible to see the residual effects of prior year teachers" (Sanders & Rivers, 1996, p. 4). The findings demonstrate that students assigned to ineffective teachers continue to show the effects of such teachers even when these students are assigned to very effective teachers in subsequent years. Although an effective teacher can facilitate excellent academic gain in students during the years in which they are assigned to them, the study found that "the residual effects of relatively ineffective teachers from prior years can be measured in subsequent student achievement scores" (Sanders & Rivers, 1996, p. 4; see also Jordan, Mendro & Weersinghe, 1997).

Furthermore, when the data were aggregated by student achievement level, it was found that ineffective teachers were ineffective with all students, regardless of the prior level of achievement. As the level of teacher effectiveness increased, students of lower achievement were the first to benefit, and only teachers of the highest effectiveness were generally effective with all students. Only the teachers in the fifth quintile produced adequate gains in the highest-achieving students. Because of this, lower-achieving students were more likely than higher-achieving students to make adequate gains, year to year.

The implications of this finding are that only the most effective teachers-the top 20 per cent-are providing instruction that produces adequate gain in high-achieving students, while students in the lower achievement levels profit from all but the least effective teachers. Therefore, the majority of the brightest students fail to achieve to their potential year after year and, in the long run, attain a level of achievement far below that of their more fortunate peers who have benefited from the most effective teachers. This effect is observable in school systems that vary extensively in socioeconomic level, racial composition, and location. Sanders and Rivers (1996, p. 6) state that

the teacher effects are both additive and cumulative with little evidence of compensatory effect of more effective teachers in later grades. The residual effects of both very effective and ineffective teachers were measurable two years later, regardless of the effectiveness of teachers in later grades.

Another finding emerged from the Sanders and Rivers study. On examination of the racial composition of the classes of teachers by effectiveness quintiles, they (p. 6) found that

more black students than would be expected, based on the ethnic makeup of the system, were assigned to the least effective teachers. . . . However, the achievement within the two ethnic groups is comparable across the five levels of teacher effectiveness. These analyses suggest that students of the same prior levels of achievement tend to respond similarly to teacher effectiveness levels.

In other words, African American students and white students with the same level of prior achievement make comparable academic progress when they are assigned to teachers of comparable effectiveness. However, at least in the system studied, black students were disproportionately assigned to the least effective teachers. Regardless of race, students who are assigned disproportionately to ineffective teachers will be severely academically handicapped relative to students with other teacher assignment patterns.

Discussion

From the research findings summarized in this report and from other studies, as well (see, for example, Jordan et al., 1997), it is clear that teacher effectiveness is the major factor influencing student academic gain. This finding holds major implications for the field of educational evaluation, the most obvious of which is that educational assessment that does not address teacher effectiveness is, at the very least, seriously limited in its ability to serve its primary purpose-to provide a basis in fact for educational improvement. It is only when educational practitioners-teachers as well as school and school system administrators-have a clear understanding of how they affect their students in the classroom that they can make informed decisions about what to change and what to maintain. Linking student outcomes to school, school system, and teacher effects can provide this invaluable information.

With the advent of TVAAS methodology, obstacles that previously hindered the effective and appropriate use of standardized test data for the purpose of educational evaluation have been removed. These data can now provide best linear unbiased predictors of the effectiveness of individual teachers, as well as schools and school systems, in regard to student academic gain. Even though standardized test data cannot provide all the information necessary to determine the effectiveness of a teacher or a school, they can be invaluable and indispensable in linking student outcomes to specific teachers and schools.

If an assessment system is to provide data at the teacher level, students must be tested annually with fresh, equivalent, nonredundant tests that exhibit a high level of reliability and validity, regardless of whether they are standardized tests or are some alternative type of outcomes assessment. This testing should continue throughout the scholastic career of each child. Currently, many states routinely test children in the lower grades, but only recently have significant numbers of states and districts begun constructing, piloting, and administering end-of-course tests in the high schools. These new tests on the secondary level are vital if academic progress is to be maximized for every student. The previously mentioned research on the variation among ACT scores of students who were in the top quartile according to their eighth-grade TCAP scores is a particularly good example of why this is so. The fact that there is great variation in academic attainment of students of similar ability depending on where they attend high school is valuable information, to be sure, but it is impossible to know from one omnibus test what went wrong in which class under which teacher that would account for the fact that some very bright students end up scoring considerably lower as seniors than their academic peers in the eighth grade do. To provide that level of diagnostic information, end-of-course data are essential.

As more and more states codify academic standards that all students are expected to meet, the question of responsibility becomes paramount. If students are responsible for attaining the standards, then teachers are responsible for teaching them. If students have differing abilities to learn, then somehow all must still be presented with the opportunity to learn. Responsible assessment is a necessary component of responsible teaching and learning.

Any educational assessment system is limited if it does not provide measures of the effects of schools, school systems, and teachers on the academic growth of students.

Without this information, educational improvement efforts cannot address the real factors that have been proven to have the greatest effect on student learning. Assessment systems of appropriate tests and methodologies for analyzing them are now available. It is through these new assessment systems that take into account the impact of educational experiences in the school and in the individual classroom that the promise of effective education for all students can come closer to realization.

References

Graphical Summary of Educational Findings from the Tennessee Value-Added Assessment System (TVAAS), 1995. (1995). Knoxville: University of Tennessee Value-Added Research and Assessment Center.

Jordan, H.R., Mendro, R.L., & Weersinghe, D. (1997). *Teacher effects on longitudinal student achievement: A preliminary report on research on teacher effectiveness*. Paper presented at the National Evaluation Institute, Indianapolis, IN.

McLean, R.A., & Sanders, W.L. (1984). *Objective component of teacher evaluation: A feasibility study*. Working Paper No. 199. Knoxville: University of Tennessee, College of Business Administration.

Sanders, W.L., & Rivers, J.C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Research Progress Report. Knoxville: University of Tennessee Value-Added Research and Assessment Center.

Sanders, W.L., Saxton, A.M., & Horn, S.P. (1997). The Tennessee Value-Added Assessment System (TVAAS): A quantitative, outcomes-based approach to educational assessment. In Millman, J. (ed.), *Grading Teachers, Grading Schools*, Thousand Oaks, CA: Corwin Press.

Sanders, W.L., Saxton, A.M., Schneider, J.F., Dearden, B.L., Wright, S. Paul, & Horn, S.P. (1994). Effects of building change on indicators of student academic growth. *Evaluation Perspectives*, 4(1) pp 3 and 7.

Wright, S.P., Horn, S.P., & Sanders, W.L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 1(1), 57-67.