**Special Studies and Evaluation**
**Office of Policy and Evaluation**
**California Department of Education**

**August, 1998**

# Research Brief

# Value-Added Methods for Accountability

"Value-added" in an educational context relates to a type of data analysis used to describe student academic gains over time for the purpose of educational accountability. A general definition is "what schools add to their pupils' knowledge, skills, and understanding between one age and another."

The appeal of the value-added methodology is that it attempts to more accurately describe the amount of student academic growth from one grade level to the next regardless of the student's academic level at the beginning of the school year. In this regard, it addresses one of the major drawbacks in educational accountability--how to fairly hold teachers and schools accountable for student learning but not for factors beyond the control of the teacher or school, e.g., socioeconomic status, parent education level, and mobility.

More specifically, the value-added methodology is a data analysis technique that compares a student's test score to the prediction of how well the student can be expected to achieve based on his or her performance during the previous year or years. An estimate of the deviation from expectation is considered to be the "value added" by the teacher or, at a more aggregate level, by the school.

**Background**

Over the past 50 years, efforts to connect student learning meaningfully to teachers or schools has had limited success. The difficulty has not been whether student learning is an important goal of teaching or the school's outcomes but rather how best, in high stakes contexts, to evaluate how well teachers and schools accomplish the task of student learning.[1]

Although cross-sectional, snapshot comparisons of aggregate student achievement from year to year are widely used methods for accountability reporting (e.g., comparisons of this year's 4th grade scores to last year's 4th grade scores), several recent new approaches have been developed which use learning gains to evaluate teachers and/or schools. These approaches attempt to address past problems of unreliability of the measurement of teacher or school effectiveness. Two of the more prominent approaches are the Tennessee Value-Added Assessment System (TVAAS) and the Dallas Value-Added Accountability System.

**Current Uses of Value-Added Methodologies**

<u>Tennessee Value-Added Assessment System (TVAAS)</u>

Implemented in 1992 as the "heart" of a comprehensive reform package passed in 1992, TVAAS was highly supported by the Tennessee Business Roundtable which desired strong educational accountability. TVAAS is used to evaluate teachers, schools, and school systems through a specific method of computing and reporting achievement gains.

Using the CTBS/4 instrument, TVAAS analyzes academic gains for grades 3 through 8 in five subject areas: mathematics, reading, language arts, social studies, and science. The scaled scores from this test are included with the record for each child along with information on where the child attends school and which teacher(s) he or she had for each grade or subject. Utilizing information from at least three and not more than five years of data, each new year's data are merged with those of previous years to give the most complete record possible. The value-added procedures compare the gains each student makes from year to year on the norm-reference part of the test. These student gains are subsequently compared to the gains made by a national normative sample for that same subject between those same grade levels. Thus, if the normal gain from 4th to 5th grade in mathematics were 15 points, a 5th grade teacher's students who averaged a 15 point gain for the year

would score "100" or 100 percent or normal gains. A teacher whose students averaged an 18 point gain would score 120, and so forth.[2] The model divides teachers into five categories--from low to high effectiveness--based on whether their pupils score better or worse than anticipated over a four-year period.

**Selected Research Findings from TVAAS Data[3]**

- The single largest factor affecting academic growth of student populations is differences in the effectiveness of individual classroom teachers.
- The cumulative and additive effects of the quality of teachers a student encounters are large. Three straight years of the most effective teachers from grades 3 through 5 result in math scores averaging at the 85th to 95th percentile. Three straight years of the least effective teachers result in scores from the 35th to 45th percentile.
- The effectiveness of a school in helping students make gains cannot be predicted based on its racial or economic make-up.
- Low achieving students are the first to benefit as teacher effectiveness improves. High achieving students do not have the opportunity to demonstrate academic growth at the same rate as low achieving students.
- More variability in teacher effectiveness exists in the higher elementary grades than in the lower elementary grades.
- In the aggregate, school principals have very little impact on the academic growth of their school population. Teachers are functioning as independent entities with little evidence of a community effect.
- When populations of students change buildings, there is a measurable drop in academic growth regardless of grade level for the first year in the new building.

TVAAS uses a mixed-model statistical methodology developed by William L. Sanders based upon a technique developed by C.R. Henderson in the area of genetics.[4] This methodology appears to resolve at least three past problems with gain score computing and reporting: (1) control for confounding variables, (2) accommodation for missing data, and (3) regression to the mean.

Control for confounding variables. Using a different approach from typical statistical regression methods, the TVAAS model filters confounding influences such as socioeconomic status and parent education without having direct measures of all of the concomitant variables. By focusing upon measures of academic gain, each student serves as his or her own "control" or, in other words, each child can be thought of as a "blocking factor" that enables the estimation of school system, school, and teacher effects upon the academic gain with the need for few, if any, of the typical background variables.

Accommodation for missing data. The method is able to use all available data but not be hindered by fractured records. The TVAAS approach can accommodate (and not "overreact" to) missing data, thus allowing for analysis of all available data regardless of high mobility as is found in some schools.

Regression to the mean. The TVAAS method is designed to provide protection against the severe misclassification of a school's, system's, or teacher's influence on student gain. It assumes all teacher (or school) effects as the average of their school system until the weight of the data pulls specific estimates away from the school system mean, thereby protecting from misclassification. This function ensures protection particularly when a teacher's score is based on small or varying quantities of student scores.

Over the past six years since its implementation, the TVAAS system has been praised by some and questioned by others. The Tennessee Comptroller of the Treasury commissioned two evaluations of the TVAAS.[5] In addition, due to the newness of the statistical procedures and the controversy surrounding use of TVAAS to evaluate teachers, other reviews of the Tennessee system by the educational statistical and policy community have been conducted.[6]

Findings From Evaluations of TVAAS. The following summarizes key findings of evaluations of TVAAS:

- TVAAS has led the way in developing a better method of analyzing longitudinal student data. The method successfully addresses three particularly difficult past problems in this type of research.
- Reviews of TVAAS from leaders in the field of statistics and measurement confirm that the mixed-model statistical methodology is sound overall. TVAAS was found to be reasonable and consistent with similar hierarchical linear modeling methods widely used in educational studies, although some refinements for improving the technical quality of the approach have been suggested.

- Although early in its implementation TVAAS was particularly criticized for its production of individual teacher scores, many of the large districts in Tennessee now appear to view the information as one of a number of a useful tools in teacher evaluation. Despite apparent success with at least some school districts in Tennessee, policy reviewers of TVAAS recommended that teachers not be included in the analysis.

- The mixed-model statistical technique is complex and not well-known to behavioral statisticians. One of the major criticisms of the system is the difficulty in explaining the technique to others. Reviewers recommended that the TVAAS administrators adequately explain their procedures or choose other ones that can be understood by those being evaluated.

- In the area of policy, reviewers recommended that the value-added analysis, being based solely on norm-referenced test scores, should not be the primary focus of the Tennessee accountability system. Other or additional measures should be used besides the norm-referenced part of the CTBS/4. The mixed-method analyses could accommodate any type of reliable linear measure of academic growth with a strong relationship to the curriculum.

- Snapshot/status score levels should be reported with gain scores. In addition, audit and verification procedures should be implemented as part of the system.

- TVAAS currently provides information on student gains related to teachers, schools, and school systems but does not provide information on what types of teaching or school practices are most or least effective.

Dallas Value-Added Accountability System

Since 1992, as an outgrowth of a decade's work, the Dallas Independent School District (DISD) has determined effective schools using a value-added accountability system. The system recently has been expanded to include the identification of effective teachers and to shape teacher evaluation for the district. The DISD accountability method uses a two-stage hierarchical linear model (HLM) methodology for student outcome data. Student achievement data include a state-mandated criterion-referenced test (Texas Assessment of Academic Skills), a norm-referenced test (Iowa Tests of Basic Skills), and a series of approximately 150 end-of-course criterion-referenced tests (Assessment of Course Performance). For the first stage, regression analysis is used to control for preexisting student differences (confounding influences) in ethnicity, gender, language proficiency, and socioeconomic status. For the second stage, the HLM methodology controls for the effects of prior achievement or attendance and school-level influences including mobility, crowding, percentage minority, and socioeconomic status. School-level outcome variables are analyzed with a simple multiple regression model using two prior years of data for a school on each variable.

Findings From Reviews of DISD. The following summarizes key findings from two recent critiques of the Dallas Value-Added Accountability System.[7]

- Although the DISD value-added system uses some controls for confounding factors (e.g., socioeconomic status, etc.), it is not considered as strong a basis for measuring growth as the TVAAS model. It appears that the Dallas method only partially addresses confounding factors and does not accommodate the other typical problems with gain scores: missing data or regression to the mean. However, reviewers suggest that the Dallas methodology may still be useful with some changes in its approach. For example, rather than pre-controlling for confounding variables using the regression approach, changes in performance should be based on the student's gain score as used in TVAAS where the student acts as his or her own control. Likewise, the measuring instruments used by Dallas should be equated in the proper manner and have the metric necessary for modeling change and growth (i.e., measures are linear on a continuous scale).

- Policy reviewers recommended that the approach be used to evaluate schools and school systems rather than teachers.

- DISD uses a statistical technique which, as with the TVAAS method, is complex and difficult to explain. Administrators of DISD should adequately explain their procedures or choose other ones that can be understood by those being evaluated.

**Summary and Future Issues**

Following is a summary of major conclusions from analyses of current uses of value-added methods and key issues for considering these types of methods for California's accountability system.

- The TVAAS statistical technique appears to be a useful and valuable tool for addressing important research questions pertaining to student growth over

time. However, as noted by Sanders and his colleagues, "schools, school systems, and teachers cannot be assessed solely on the basis of TVAAS." The system should use the most instructionally sensitive assessment devices possible, and the norm-referenced tests should be supplemented with criterion-referenced tests well-aligned to a given school's district-sanctioned or state-sanctioned curriculum.

- Value-added analysis for teacher evaluation was generally not recommended by policy reviewers. Rather, as suggested by one reviewer, it may be more worthwhile to strengthen current teacher evaluation processes and hold the site principal responsible for conducting such evaluations.

- Reviewers of TVAAS and the Dallas system overwhelmingly stated the necessity for clear and understandable reporting. For any accountability system, statistical methods must be selected that are at least intuitively comprehensible to those who are evaluated, and they must be regarded by educators and policymakers as sensible ways in which to level the evaluative playing field. Value-added methodologies and results, which involve particularly sophisticated and complex issues, must be introduced and reported in a language that can be understood by those being evaluated. Both value-added and snapshot/status score levels should be reported at the same time. Snapshot/status score levels should be the focus of school and system scores.

- If value-added methods were to be used in California, a statewide student information system that includes student scores on successive grade-level assessments would need to be available. Student scores would need to be provided on a common scale, a single developmental scale of measurement. Statistical, measurement, and large database expertise would be needed to design, implement, and manage information processing and reporting. Procedures would need to be established whereby value-added calculations are audited and verified. Security and privacy processes and procedures would need meticulous attention, particularly if the value-added approach were used to evaluate teachers.

- As California's accountability system evolves, value-added methodologies have potential for supplementing and improving traditional snapshot information about student performance. Further investigation of these value-added techniques is needed to determine whether such methods hold promise for California's unique needs and accountability goals.

This Research Brief was prepared by Jan Volkoff of the Office of Policy and Evaluation. For more information, contact Jan Volkoff at (916) 657-4282 or jvolkoff@cde.ca.gov. This Research Brief may also be found on the Internet at <www.cde.ca.gov/ftpbranch/ope/sse>.

---

[1] Millman, J., & Schalock, D.H. (1997). Beginnings and introduction. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* Thousand Oaks, CA: Corwin Press, Inc.

[2] Sanders, W.L. (1998). *Welcome to the site for comprehensive information on the Tennessee value-added assessment system* (on-line). Available: www.tvaas.com

[3] Pipho, C. (1998). The value-added side of standards. *Phi Delta Kappan, 79*(5), 341-342.
Sanders, W.L. (1998). *Welcome to the site for comprehensive information on the Tennessee value-added assessment system* (on-line). Available: www.tvaas.com

[4] Sanders, W.L. (1989). A multivariate mixed model. In *Applications of mixed models in agriculture and related disciplines*. (Southern Cooperative Series Bulletin No. 343, pp. 138-144). Baton Rouge: Louisiana Agricultural Experiment Station.

[5] Baker, A.P., & Xu, D.. (1995). *The measure of education: A review of the Tennessee value added assessment system*. Nashville, TN: Comptroller of the Treasury, Office of Educational Accountability.
Bock, R.D., Wolfe, R., & Fisher, T.H. (1996). *A review and analysis of the Tennessee value-added assessment system*. Nashville, TN: Comptroller of the Treasury, Office of Educational Accountability.

[6] Shinkfield, A.J., & Stufflebeam, D. (1995). *Teacher evaluation: Guide to effective practice*. Norwell, MA: Kluwer Academic Publishers.
Millman, J. (1997). *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* Thousand Oaks, CA: Corwin Press, Inc.

[7] Thum, Y.M., & Bryk, A.S. (1997). Value-added productivity indicators: The Dallas system. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* Thousand Oaks, CA: Corwin Press, Inc.
Sykes, G. (1997). On trial: The Dallas value-added accountability system. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* Thousand Oaks, CA: Corwin Press, Inc.