Journal of Personnel Evaluation in Education 8:299–311, 1994

The Tennessee Value-Added Assessment System (TVAAS): Mixed-Model Methodology in Educational Assessment

WILLIAM L. SANDERS
Professor and Director, University of Tennessee Value-Added Research and Assessment Center, University of Tennessee, P.O. Box 1071, Knoxville, TN 37901-1071

SANDRA P. HORN
Educational Consultant, University of Tennessee Value-Added Research and Assessment Center.

Introduction

Background

Over the past decade, Tennessee, like so many other states, has continuously sought to improve educational opportunities for its students. The first wave of reform resulted in the Comprehensive Education Reform Act of 1983 (CERA). CERA created a Career Ladder Program (a merit pay system for teachers) and a Basic Skills Program. CERA also led to the articulation of grade and subject curricula and the development of curricular frameworks for the state of Tennessee.

At about this time, independent of the efforts of the Tennessee Department of Education, two statisticians, Dr. William L. Sanders and Dr. Robert A. McLean of the University of Tennessee, had begun to explore the feasibility of using a statistical mixed-model methodology to eliminate many of the previously cited impediments to incorporating student achievement data in an educational outcome-based assessment system. These problems include but are not limited to the following: missing student records, various modes of teaching (self-contained classroom versus departmentalized instruction versus team teaching), teachers changing assignments over years, transient students, regression to the mean, different variance-covariance structures across school systems, and the need to include concomitant covariables as needed. A decade of work has demonstrated that a system can be developed to eliminate, or at least trivialize, these problems.

In 1984, McLean and Sanders published a working paper on the use of student achievement data as a basis for teacher assessment (McLean & Sanders, 1984). Utilizing three years of gain scores from Knox County students' performance on the California Achievement Test in grades 2 through 5, Sanders and McLean developed a statistical system of analysis based on Henderson's mixed-model methodology. This study rendered the following findings:

1. There were measurable differences among schools and teachers with regard to their effect on indicators of student learning.
2. The estimates of school and teacher effects tended to be consistent from year to year.
3. Teacher effects were not site specific; a gain score could not be predicted by simply knowing the location of the school.
4. There was very strong correlation between teacher effects as determined by the data and subjective evaluations by supervisors.
5. Student gains were not related to the ability or achievement levels of the students when they entered the classroom.

Subsequent studies incorporating data from Blount County and Chattanooga City Schools bore out the initial findings. The study of the Chattanooga City Schools, a system that includes many inner-city schools, produced a new finding not evident from the previous studies of systems that were primarily suburban and rural: the estimate of school effects was not related to the racial composition of the student body.

Even though these findings indicated the efficacy and utility of this assessment approach, the Sanders model (as this process has been labeled in Tennessee) was for several years thereafter known only to a small circle of educators and statisticians.

In 1988, educational reform in the state took a different direction. The Tennessee Department of Education developed a document titled *21st Century Challenge: State Goals and Objectives for Educational Excellence* in response to the *America 2000 Program*; the Tennessee State Board of Education put forth its *Master Plan for Tennessee Schools*; and the Tennessee Higher Education Commission developed *Tennessee Challenge 2000* for postsecondary educational institutions. The goals and objectives of these governing bodies were coordinated to form an educational framework that would address learner needs and expectations from preschool through adulthood. At every level, the need for accountability and assessment was recognized as an essential component of educational improvement.

When the recommendations of the governing educational bodies were submitted to the Tennessee General Assembly for legislative action in the form of the Education Improvement Act, it became necessary to specify the means by which teachers, schools, and school systems would be held accountable for meeting the goals and objectives set forth for Tennessee's educational systems. Since the focus of the accountability movement was on the *product* of the educational experience rather than the *process* by which it was to be achieved, the outcomes-based assessment system Sanders and McLean had been refining was an obvious choice for consideration. In 1991 when the Education Improvement Act was adopted, the model now known as the Tennessee value-added assessment system (TVAAS) formed an integral part of the legislation.

Philosophical underpinnings of the Tennessee value-added assessment system

Ralph W. Tyler, a major force behind the development of modern educational evaluation, proposed that evaluation should be a process of comparison between stated objectives and actual outcomes. In Tennessee, the connection between objectives and outcomes is explicitly recognized. *The Master Plan for Tennessee Schools 1993* (Tennessee State Board of Education, 1992) sets forth goals in eight key result areas: early childhood education, primary and middle-grades education, high school education, technology, professional development and teacher education, accountability, school leadership and school-based decision making, and funding. The goal for the accountability component of the master plan is as follows: "State and local education policies will be focused on results; Tennessee will have assessment and management information systems that provide information on students, schools, and school systems to improve learning and assist policy making" (p. 7). Here, Tyler's conception of evaluation is readily discerned. Assessment is recognized as a tool for educational improvement, providing information that allows educators to determine which practices result in desired outcomes and which do not. By focusing on outcomes rather than the processes by which they are achieved, teachers and schools are free to use whatever methods prove practical in achieving student academic progress. Value-added assessment is one means recognized by the state of Tennessee for assessing progress toward the academic goals set forth in the master plan (p. 17) and the Education Improvement Act.

Astin (1982, p. 14) states that "the basic argument underlying the value-added approach is that true excellence resides in the ability of the school or college *to affect its students favorably*, to enhance their intellectual development, and to make a positive difference in their lives." TVAAS was developed on the premise that society has a right to expect that schools will provide students with the opportunity for academic gain regardless of the level at which the students enter the educational venue. In other words, all students can and should learn commensurate with their abilities. By focusing on the gains that all students make from year to year, the school systems and the individual schools deemed to be most effective by TVAAS are those that provide educational opportunities for all learners—the advanced learner as well as the slower learner.

A description of the Tennessee value-added assessment system

General information

TVAAS is a statistical process that provides measures of the influence that school systems, schools, and teachers have on indicators of student learning. Initially, TVAAS will furnish this information on the system level for each school system in Tennessee for grades 3 through 8 in math, science, reading, language, and social studies by using the scale scores from the Tennessee Comprehensive Assessment Program (TCAP). TVAAS will be extended to cover grades 9 through 12 when subject-matter-specific tests that can provide comparable data for these grades have been developed and validated—by law, no later than July 1, 1999. TVAAS is mandated by the Education Improvement Act, which took effect July 1, 1992.

TVAAS analyzes the scale scores students make on the norm-referenced items of the TCAP. The pattern of the scale scores over the child's school career forms a profile of academic growth. A database containing the merged records of all students in Tennessee who have taken the TCAP tests during the past four years has been constructed. At present, it contains more than 3.3 million student records. This number will continue to grow over time and will enable continued tracking of the academic growth of each student.

The Education Improvement Act (EIA) mandated that school system effects on the educational progress of students for grades three through eight, as determined through the use of TVAAS, be reported for systems statewide no later than April 1, 1993. These reports have been distributed to each school system in Tennessee. They have also been released to the public and will be updated annually.

The EIA set July 1, 1994 as the deadline for issuing the first set of reports on individual school effects. This set of reports is also be available to the public and will be revised on a yearly basis.

The individual teacher effects for teachers of grades 3 through 8 are to be reported to the teacher, appropriate administrators, and school board members no later than July 1, 1995, according to the EIA. These reports relating to the influence of individual teachers on the rate of student learning will not be available to the public. Reports on all levels will be based on at least three years of data and no more than five years of data.

The assessment of schools and school systems

The assessment of schools and systems, although it requires massive computing capabilities, logistical planning, and statewide testing, is fairly simply explained. The mixed-model equations incorporate the scale scores of all the students taking the norm-referenced portion of the TCAP in all five subjects, modeling a learning profile of each student for each subject as explained in the section above. These profiles are grouped by system or school, as the case may be. The gain scores of a school's or system's students are estimated and are then compared to the national norms. Deviation from the norm gain is reported for each subject and grade. The school or school system can then identify where students are achieving normally, outstandingly, and substandardly.

The state of Tennessee monitors the gains of all school systems in the state for subjects or grades that are not achieving national norm gains. Those systems achieving two or more standard errors below the national norms must show positive progress or risk intervention by the state. Each school and system is expected to achieve the national norm gains regardless of whether its scale scores are above or below the national norm.

Assessment of teachers

The assessment of teachers is generally the most controversial aspect of any educational evaluation system. The great variety of teaching situations and the endless diversity of the student population have rendered each attempt at teacher assessment suspect to a greater or lesser degree. TVAAS is not the first to base teacher assessment on student achievement. However, important differences exist between TVAAS and its predecessors.

Beginning with the 1992–1993 school year, detailed information identifying each teacher with the students he or she teaches will be collected annually. Included in the data will be subjects taught to each student and the proportion of time each student spends with a teacher. If team teaching or departmentalized teaching takes place, it will be identified along with the proportion of each subject the teacher is responsible for teaching. From attendance records submitted to the state, it will be determined whether each student has been present in each teacher's class the required 150 days in a given school year because students who have not been in a teacher's class at least 150 days in a year will not figure into the teacher's assessment. By 1995 when the teacher assessments are scheduled for delivery, three years of such data will be available. The EIA requires that teacher assessment be based on at least three and no more than five years of data.

Test reliability and relevance

TVAAS uses scale scores from the norm-referenced items on the Tennessee Comprehensive Assessment Program (TCAP), which was first implemented in the 1989–1990 school year. The norm-referenced part of TCAP, the CTBS/4, is a nationally normed test mandated in Tennessee for grades two through eight and grade ten. It assesses skills in reading, language arts, math, science, and social studies. The norms for the test were established in 1989. Williams (1989) states in his review of customized standardized tests that, in Tennessee, "the norm-referenced module was specifically created so that it has proper statistical characteristics of reliability, adequate floors and ceilings, and articulation across test levels." To ensure test validity, the EIA mandates that "fresh, non-redundant tests" be used each year. This means that only a small percentage of the items on the CTBS/4 can be carried over from one year to the next. Moreover, rigorous sanctions are provided in the EIA for any breach of test security. The relevance of the test to Tennessee's academic program may be inferred from the tendency of scores across the state to approximate or slightly exceed the national norms in all subject areas and all grades.

The scores from the CTBS/4 cannot reflect the totality of a student's learning experience or progress. However, these scores, as they are utilized by TVAAS, provide an unbiased estimate of the influence of school systems, schools, and teachers on students' academic growth in the subjects tested. This academic growth is and should be a primary goal of Tennessee's educational system. TVAAS uses data from a testing system already mandated and in place statewide. However, should better tests be developed in the future, no major alterations would have to be made in order for TVAAS to incorporate new sources of data, as long as the methods of assessment can provide linear metrics.

Problems of using student achievement data in educational assessment

The use of student achievement data to directly measure educational outcome has much intuitive appeal and has been advocated by many. However, serious proponents of this approach have recognized several difficulties that must be overcome in order to insure a fair and reliable system for outcome assessment.

These difficulties may be categorized into problems associated with (1) the definition and construction of appropriate metrics and (2) development and implementation of a statistical methodology that will allow fair and unbiased assessment of school systems, schools, and teachers when nonrandom assignment of students is assumed. (We will not deal with the definition and construction of metrics but rather will assume that metrics exist or can be constructed that adequately proxy learning.)

Even when metrics exist with suitable characteristics, many problems of school and teacher assessment remain. The ensuing discussion will focus on the problems associated with the estimation of the influence of *teachers* on the rate of student gain because at the classroom level the problems are more difficult than at the school or school system level.

Since random assignment of students to teachers is usually not practiced and seldom is possible, simple means of class achievement test scores are seriously biased by many factors other than teacher influences that affect student learning. Travers (1981) listed (1) teacher influences, (2) parental influences, (3) genetic endowment, (4) other school influences, and (5) availability of materials as being some of the most important factors that determine the rate of student learning.

Later, in their attempt to develop a value-added method of evaluation based on student test scores, Bingham, Heywood, and White (1991) list forty-four variables under five major categories—individual characteristics, family characteristics, classroom characteristics, school characteristics, and academic performance—which they determined were independent of the input of school and teacher for the subject school system during the years of their study.

In spite of the detailed character of this listing, Bingham et al. (1991, pp. 200–201) point out that these variables may be pertinent only to the particular school system they studied and perhaps only during the years in which the research took place. Obviously, any system that will fairly and reliably assess the influence of teachers on student learning must partition teacher effects from these and other factors. However, it is a hopeless impossibility for any school system to have all the data for each child in appropriate form to filter all of these confounding influences via traditional statistical analysis.

Using a different approach, the three studies conducted by Sanders indicate that these influences can be filtered without having to have direct measures of all of the concomitant variables. By focusing on 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 on the academic gain with the need for few, if any, of the exogenous variables.

In an attempt to partition the teacher and school effects from the partial confounding with class ability level, the well-known linear model techniques of analysis of covariance and ordinary multiple regression have been suggested by Millman (1981) and others. The obvious intent was to adjust differences that exist among students to enable a fairer evaluation of teachers. However, if these simple approaches are applied, and even if all of the concomitant data were available, still unanswered is the well-known problem of regression to the mean of the teacher effects that would provide unfair rankings of teachers with varying quantities of student achievement records. Also, the problem of missing student records due to transient student populations, students being absent during the time of testing, and so on would result in very few usable records if these traditional methods were employed.

Advantages of considering educational outcome assessment from student data as a statistical mixed-model problem

Traditional multiple regression or analysis of covariance can be characterized as techniques in linear model analysis with all fixed variables. If the problem is viewed not as a fixed-effects problem but rather as a mixed-model problem with both fixed and random effects, then much established theory and methodology exist that offer solutions to many of the problems that have been cited as reasons for not doing educational outcome assessment from student achievement data.

General form of Henderson's mixed-model equations (MME)

$$y = XB + ZU + e,$$

where,

- y in the context of teacher evaluation is the $m \times 1$ observation vector representing all of the scale scores for individual students for all academic subjects tested over all grades.
- X is a known $m \times p$ matrix.
- B is an unknown $p \times 1$ vector of fixed effects.
- Z is an $m \times q$ incidence matrix.
- U is an unobservable $q \times 1$ random vector.
- e is an $m \times 1$ vector with $E(e) = 0$.

Both U and e have null means and variance:

$$Var\begin{bmatrix} U \\ e \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$$

G and R are known and nonsingular. R is the variance-covariance matrix that reflects the correlation among student scores within teacher. G is the variance-covariance matrix that reflects the correlation among teacher effects (both R and G are assumed block diagonal in the context of teacher evaluation). If (U, e) are normally distributed, the joint density of (y, U) is maximized for variations in B and U by the solution to the following equations:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z+G^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

Let a generalized inverse of the coefficient matrix be

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z+G^{-1} \end{bmatrix}^{-} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

Some of the properties of a solution of these equations are as follows (Henderson, 1984):

1. $K'b$ is best linear unbiased estimate (BLUE) of the set of estimable linear functions, $K'B$.

2. u is the best linear unbiased predictor (BLUP) of U.

   a. $E(U \mid u) = u$.
   b. $\text{var}(u - U) = C_{22}$.
   c. $\text{var}(K'b + M'u - K'B - M'U) = (K'M')C(K'M')'$.
   d. u is unique regardless of the rank of the coefficient matrix.

3. $K'b + M'u$ is BLUP of $K'B + M'U$ provided $K'B$ is estimable.

4. With G and R known, the solution is equivalent to generalized least squares and if u and e are multivariate normal then the solution is maximum likelihood.

5. If G and R are not known, then as an estimated G and R approach the true G and R, the solution approaches the maximum likelihood solution.

6. If u and e are not multivariate normal, then the solution to the MME still provides the maximum correlation between U and u.

For an Introduction to Henderson's mixed-model methodology, see McLean, Sanders, and Stroup (1991).

Why should teacher effects be considered random instead of fixed?

Historically, classification variables in a linear model context that have their own probability distribution have been referred to as random effects. Since in the context of teacher evaluation other variables that do not have their own distribution (fixed effects) sometimes may be included to ensure fair evaluation, it is often more reasonable to view teacher evaluation as a mixed-model problem. When this is the case, solutions to Henderson's mixed-model equations (MME) provide BLUP of the random effects while providing opportunity for the inclusion of both continuous and classification fixed effects. This is a sufficient procedure to provide the flexibility necessary to handle the diversity of models which could be encountered in teacher assessment. Additionally, since BLUP is a "shrinkage" estimate of the realized value of the random variable (Harville, 1976), then BLUP is a solution to the regression to the mean problem, which has been long recognized as an impediment to the use of student data in an assessment system for teaching effectiveness.

Concept of best linear unbiased prediction (BLUP)

To illustrate the concept of best linear unbiased prediction, a restatement of an example presented by Henderson (1973) from Mood (1950) is presented.

Given that the population mean and variance of true IQ is 100 and 225, respectively, and if an individual takes one IQ test and scores 130 on a test that has test error variance of 25, what is the best prediction of true IQ of that individual? In this example,

$$\begin{matrix} \textit{Prediction} \\ \textit{of true IQ} \end{matrix} = \begin{matrix} \textit{Pop} \\ \textit{Mean} \end{matrix} + \left(\begin{matrix} \textit{Mean of} \\ \textit{IQ test} \end{matrix} - \begin{matrix} \textit{Pop} \\ \textit{Mean} \end{matrix} \right) \times \left[\frac{\textit{Var (Pop.)}}{\textit{Var (Pop.)} + \left(\frac{\textit{Var (Test)}}{\textit{No. of tests}} \right)} \right]$$

$$= \quad 100 + (130 - 100) \times \left[\frac{225}{225 + \left(\frac{25}{1} \right)} \right] = 127.$$

The best prediction of this individual's true IQ is not 130 but rather 127. Why is this so? This expression for the conditional mean of true IQ given IQ test score may be obtained from the joint distribution of true IQ and IQ test score if both true IQ and the errors of the test are assumed to be normally distributed (Searle, 1971, p. 461). Note that this prediction of true IQ is pulled ever closer to the population mean as the ratio of test error variance to population variance increases or as N becomes smaller. Thus, if a little information is available, a prediction close to the population mean tends to be best. If more information is available, a prediction closer to the sample mean is best.

This pulling of the prediction closer to the mean as a function of distance, ratio of the variances, and quantity of information is the essence of the BLUP concept. The concept of BLUP offers an explanation of and a solution to the regression-to-the-mean problem.

The problem of missing data

In the original Knox County study, the gain in scale score points for each student was calculated for each student and was used as the response variable in the mixed model equations. This rather simplistic approach was sufficient to establish the feasibility of the methodology. However, to calculate the gain for each student over multiple years requires no missing data for all year-academic-subject combinations. This requirement ensures the undesirable result that only a small fraction of student outcomes will be included in an assessment process.

In later work using mixed-model methodology, it has been found that complete information for each student is not necessary to provide estimates of the influence of teachers on the gain of a population of students.

Consider the following model to be applied to the data from one specific school system:

$$Y(ijkl) = Mu(ij) + \text{year*subject*teacher}(ijk) + e(ijkl)$$

where,

$Y(ijkl)$ =	the student record for the ith year, the jth academic subject, the kth teacher and the lth student record within year, subject, teacher.
$Mu(ij)$ =	the population mean with the ith year and the jth subject.
year*subject*teacher(ijk) =	the kth teacher in the ith year and the jth subject.
$e(ijkl)$ =	deviation of the lth student score around the year*subject*teacher(ijk) subgroup mean.

Now, consider $Mu(ij)$ to be fixed, and the year*subject*teacher(ijk) to be random with G as the variance-covariance matrix among the year*subject*teacher combinations. Let R be the variance covariance matrix among student records within year*subject*teacher.

The fixed portion of the solution to these mixed-model equations will contain the estimated means for the year*subject combinations. The random portion of the solution to these equations will contain BLUP for each year*subject*teacher combination. Directly from this part of the solution vector is available a profile among teachers for each subject each year. These numbers reflect relative differences but are not at this point interpretable as gains. However, these numbers can be scaled to directly reflect gains. The third property (see the section "General form of Henderson's mixed-model equations (MME)," above) of a solution of the mixed-model equations is (3) $K'b + M'U$ is BLUP of $K'B + M'U$ provided $K'B$ is estimable. Thus, by choosing K and M appropriately, then BLUP for each teacher's gain is available with its standard error (property 2c).

Property (3) of the solution offers another powerful advantage. By choosing K and M, teachers can be profiled as math teachers, as reading teachers, as language teachers, etc. or all subjects can be combined to form an overall profile merely by changing K and M.

Another powerful advantage to this approach is that many different modes of classroom instruction can be accommodated by assigning the teacher of record to each student record within the Z-matrix. It does not matter if a child is in a self-contained classroom, a departmentalized school, or in a team-teaching situation. If the Z-matrix is encoded properly, then BLUP is provided for each teacher. Also, if teachers have assignments over grades each year (one section of fourth-grade math and three sections of fifth-grade math), then all information contributes to BLUP. This is also true for teachers changing assignments over years.

Summary

The Tennessee Value-Added Assessment System circumvents many of the problems associated with the use of student achievement data in assessment of school systems, schools, and teachers by relying on the scale scores that indicate gains students make from year to year, regardless of the point at which the student enters the classroom. Three previous studies indicate that the influence of teachers and schools on the rate of student gain are independent of the confounding of socioeconomic factors. The reports of Tennessee school system effects released April 1, 1993 confirm the earlier findings in that the school system cumulative gains for each of the five subjects were uncorrelated with the percent of students receiving free and reduced-cost meals within the system. Also, the cumulative gains for all subjects were found to be uncorrelated with the racial composition of the student body within school systems. Even so, it may be that some socioeconomic confoundings could surface in the future that would necessitate the inclusion of appropriate covariables in the mixed model equations. Current findings suggest that the number of needed covariables will be relatively small, if any; however, TVAAS readily accommodates such inclusion.

The mixed-model methodology on which TVAAS relies addresses major problems in using student achievement data in educational assessment. Among these are missing student data, diversity of teaching modes, and the regression to the mean problem. The regression to the mean question is dealt with using the concept of best linear unbiased predictor. The problems of missing student data and diversity of teaching modes are alleviated by retaining the five most current years of data for students and teachers to be included in the mixed-model process (Sanders, 1989). By using all of this information for each child and by fitting all the data from teachers over subjects and grades simultaneously, considerable robustness is achieved. This robustness has been confirmed using computer simulations to evaluate worst-case scenarios.

To fit these models to the student data for each school system within a state necessitates monumental computing efforts. For TVAAS to accomplish this task, it has been necessary to develop a software system to contend with the simultaneous computation of tens of thousands of equations. Each year, as new data are added to the system, solutions to the mixed-model equations are newly obtained. Dr. Arnold M. Saxton, Dr. Boyd L. Dearden, Mr. John F. Schneider, and Mr. S. Paul Wright have worked as a team to develop the software and hardware configurations to complete the computations. This team has also developed the reports that were distributed to Tennessee's school systems and has begun analysis of gain patterns that have emerged from the data.

Even though the first reports were issued only a year and a half ago, many educators have already acknowledged the diagnostic value of the data they have received. It is perhaps here that the impact of TVAAS will be felt most fully. The vast database is yielding far more than assessment data. Because it encompasses so much student data, educational findings that were invisible in the past are now readily apparent. For instance, it was noted that there was a dip in scores in the sixth and seventh grades across the state. When the data for homogeneous systems—those systems where all students changed schools in the sixth grade and those systems where all students changed schools in the seventh grade—were aggregated, it was found that gain scores dropped dramatically the year following the school change. The analysis of school change in systems where a variety of configurations exists confirmed the preliminary findings (Sanders, et al., 1994). Further research is needed to explore the causes of the school change phenomenon.

Many other patterns are emerging that bear investigation. Future areas of exploration may included the effects of teaching mode (cooperative learning, whole language, team teaching), class size, textbook adoptions, funding, technology, curricular innovations, and many other factors.

TVAAS offers insight and perspective in the pursuit of educational improvement. It provides a solid basis from which change can be rationally undertaken. The academic gains our students make is the measure of our success as educators as well as theirs.

Reference

Astin, A.W. (1982), Excellence and equity in American education. Washington, DC: National Commission on Excellence in Education. ERIC Document Reproduction Service No. ED 227 098.

Bingham, R.D., Heywood, J.S., & White, S.B. (1991). Evaluating schools and teachers based on student performance. *Evaluation Review*, 15, 191–218.

Harville, D.A. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. *Annals of Statistics*, 4, 384–395.

Henderson, C.R. (1973). Sire evaluation and genetic trends. *Proceedings of the Animal Breeding and Genetic Symposium in Honor of Dr. Jay L. Lush*. Champaign, IL: ASAS and ADSA, pp. 10–41.

———. (1984). *Applications of linear models in animal breeding*. Guelph, Canada: University of Guelph.

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.

McLean, R.A., Sanders, William L., & Stroup, Walter W. (1991). A unified approach to mixed linear models, *American Statistician*, 45, 54–64.

Millman, J. (ed.). (1981). *Handbook of teacher evaluation*. Beverly Hills: Sage.

Mood, A.M. (1950). *Introduction to the theory of statistics*. New York: McGraw-Hill.

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

Sanders, W.L., et al. (1994). Effects of building change on indicators of student academic growth, *Evaluation Perspectives*, 4, 3–7.

Searle, S.R. (1971). *Linear models*. New York: Wiley.

Tennessee State Board of Education. (1992). *The Master Plan for Tennessee Schools 1993*. Nashville: Author.

Travers, R.M.W. (1981). In J. Millmen (ed.), *Handbook of teacher evaluation*. Beverly Hills: Sage.

Williams, P.L. (1989). Using customized standardized tests. Contract No. R-88-062003. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education. ERIC Digest No. ED 314429.