

A Review of the Tennessee Value-Added Assessment System (TVAAS)

David A. Harville
Professor of Statistics
Iowa State University
June 6, 1995

General Comments

TVAAS is a system for estimating the influence of individual teachers, schools, and school systems on the educational progress of their students. It was developed by combining (in a rather ingenious way) a very powerful statistical methodology, known as mixed model methodology, with the concept that, during each year, the effect of each teacher (or school or school system) is to add incrementally to each student's ability in one or more subjects. TVAAS provides a sound and objective basis for comparing the additions made by each teacher (or school or school system) to their students' abilities with those made by their peers or with national norms. Its use serves to identify teachers, schools, and school systems that are seriously deficient.

For each teacher (or school or school system), TVAAS can provide an accurate and unbiased estimate of the average annual addition they have made to their students' abilities in each of their subjects. It can also provide a standard error (a measure of the precision of the estimate). The nature of the estimates, together with the availability of standard errors, insures that no teacher (or school or school system) will be "targeted" unless there is very substantial evidence of serious deficiencies. TVAAS has significant advantages over other statistical methodologies that have been proposed for assessing the influence of individual teachers, schools, or school systems. In particular, systems that (unlike TVAAS) attempt to account for student differences by using concomitant information tend to be plagued with problems occasioned by missing information. TVAAS accounts for such differences in a more direct and more effective way.

In the development of TVAAS, a number of choices had to be made with regard to the underlying model, model assumptions, etc. These choices appear to have been made in a very competent and appropriate way. The overriding considerations were accuracy and fairness.

TVAAS is a relatively sophisticated and complex system. It is computationally intensive and requires powerful computer facilities for its implementation. Great care has been exercised in developing the requisite software and procedures, and the software and procedures have undergone extensive testing.

In summary, TVAAS appears to be a statistically sound and appropriate system for estimating the influence of individual teachers, schools, and school systems. A great deal of

time, effort, and thought have gone into its development, and the assumptions underlying every "new wrinkle" have been validated for a variety of real data sets. While TVAAS itself is relatively new, the statistical methodology that underlies TVAAS has been successfully adapted for use in many areas of application — refer to Robinson's 1991 review (*Statistical Science*, vol. 6, pp. 15-51). I anticipate that this methodology will be just as successful in its application to the evaluation of individual teachers, schools, and school systems as it has been in its application to problems in other areas.

Some Specific Comments

1. *Quality control.* Adequate safeguards have been taken to insure that the assessments provided by TVAAS are essentially error-free. Much merging of test scores and much other data preparation must be carried out before the test scores can be processed by TVAAS. In developing procedures for accomplishing these preliminary tasks, great care has been taken to avoid misidentification of students or teachers.

2. *Estimation of variances and covariances.* It is assumed that a student's test scores (corresponding to as many as 5 years and 5 subjects) are correlated. The variances and covariances of these scores must be estimated. These estimates are required as input to TVAAS; TVAAS exploits the information in these estimates in estimating the influence of individual teachers, schools, and school systems. Separate estimates are obtained for each school system. For purposes of estimating the influence of individual teachers, the variance among teachers must also be estimated. Two approaches to the estimation of the variances and covariances have been devised: the more sophisticated (and more computationally intensive) of the two approaches is known as REML (for restricted or residual maximum likelihood); the second approach is simpler but possibly less accurate. REML is to be used when it is computationally feasible to do so. The methods that are being used for estimating variances and covariances are adaptations of widely used and widely accepted methodologies and can be expected to produce very satisfactory estimates.

3. *Teacher evaluation.* In the case of teacher evaluation, TVAAS is based on a linear statistical model that considers a student's test score to be the sum of the following three quantities: a parameter that is specific to year, grade, and subject, a sum of teacher effects (that are specific to year, grade, and subject), and a residual effect. The teacher effects are those for the teachers that the student had for that subject during the year of the test and during previous years. Appropriate modifications are introduced to account for team teaching and for any other deviations from standard practice. The teacher effects are modeled as realizations of random variables; and as a consequence the estimates of the teacher effects are what are known to statisticians as shrinkage estimates — this insures that

no teacher will be identified as significantly worse than other teachers in the same school system unless the evidence for such a difference is very substantial. In short, the model provides a very appropriate basis for teacher evaluation and has produced very sensible results when applied to a wide variety of data sets.

4. *Variability of results.* A report released recently (in April 1995) by the Comptroller of the Treasury of the State of Tennessee (entitled "The Measure of Education: A Review of the Tennessee Value Added Assessment System") was critical of TVAAS. In particular, the report raised questions about the relatively large yearly differences in the value-added scores for certain school systems and about the rather large differences in the value-added scores for two seemingly similar schools. The apparent implication was that these differences reflect poorly on TVAAS and raise questions about the assumptions that underlie it. However, in the example (cited in the reports) of yearly differences, it appears that TVAAS was accurately reflecting the information in the data. That is, the differences are "real," even though the reasons for the differences may not be obvious. Moreover, in the example (cited in the report) of school-to-school differences, the authors of the report made a basic error in determining the standard error of a difference of two estimates — the variance of the difference (which is the square of the standard error) should be the sum of the variances of the two estimates, not the larger of the two variances.

