

# SAS® EVAAS

## Methodology Overview

### Predictive Model

#### Overview

The predictive model is a regression-based value-added model where growth is analyzed by comparing students' expected scores or predicted achievement and their actual scores. Predicted achievement is met when students made the same amount of growth as students with the average experience in the state in that year for that assessment given their prior testing history. This model is used for the purposes of the Teacher Merit Award (TMA) program.

The model defines expected growth based on the empirical student testing data; in other words, the model does not assume a particular amount of growth or assign predicted achievement in advance of the assessment being taken by students. The predictive model defines predicted achievement within each year.

#### Technical Description

In the predictive model, each student receives an expected score or scale score representing predicted achievement based on their own prior testing history. In practical terms, this score represents the student's entering achievement because it is based on all prior testing information to date. The expected scores can be compared to the students' actual scores.

The approach is described briefly below with more details following.

- The predicted score serves as the response variable ( $y$ , the dependent variable).
- The covariates ( $x$  terms, predictor variables, explanatory variables, independent variables) are scores on tests the student has already taken.

Algebraically, the model can be represented as follows for the  $i^{th}$  student.

$$y_i = \mu_y + \beta_1(x_{i1} - \mu_1) + \beta_2(x_{i2} - \mu_2) + \dots + \epsilon_i \quad (1)$$

The  $\mu$  terms are means for the response and the predictor variables. The  $\beta$  terms are regression coefficients. Predictions to the response variable are made by using this equation with estimates for the unknown parameters ( $\mu$  terms and  $\beta$  terms). The parameter estimates (denoted with "hats," e.g.,  $\hat{\mu}$ ,  $\hat{\beta}$ ) are obtained using all students that have an actual value for the specific response and have three predictor scores. The resulting prediction equation for the  $i^{th}$  student is as follows:

$$\hat{y}_i = \hat{\mu}_y + \hat{\beta}_1(x_{i1} - \hat{\mu}_1) + \hat{\beta}_2(x_{i2} - \hat{\mu}_2) + \dots \quad (2)$$

One difficulty that must be addressed to implement the predictive model is that not all students will have the same set of predictor variables due to missing test scores. The strategy for dealing with missing

---

predictors is to estimate the joint covariance matrix (call it  $C$ ) of the response and the predictors. Let  $C$  be partitioned into response ( $y$ ) and predictor ( $x$ ) partitions, that is,

$$C = \begin{bmatrix} c_{yy} & c_{yx} \\ c_{xy} & c_{xx} \end{bmatrix} \quad (3)$$

This matrix is estimated using the Expectation Maximization (EM) algorithm for estimating covariance matrices in the presence of missing data provided by the Multiple Imputation procedure in SAS/STAT® (although no imputation is actually used). Only students who had a test score for the response variable in the most recent year and who had the required number of variables are included in the estimation. Given such a matrix, the vector of estimated regression coefficients for the prediction equation (2) can be obtained as:

$$\hat{\beta} = C_{xx}^{-1}c_{xy} \quad (4)$$

This allows the use of whichever predictors a student has to get that student's expected  $y$ -value ( $\hat{y}_i$ ). Specifically, the  $C_{xx}$  matrix used to obtain the regression coefficients *for a particular student* is a subset of the overall  $C$  matrix that corresponds to the set of predictors for which this student has scores.

The prediction equation also requires estimated mean scores for the response and for each predictor (the  $\hat{\mu}$  terms in the prediction equation). Once the parameter estimates for the prediction equation have been obtained, predictions can be made for any student with any set of predictor values, so long as that student meets the minimum requirement for number of prior test scores.

$$\hat{y}_i = \hat{\mu}_y + \hat{\beta}_1(x_{i1} - \hat{\mu}_1) + \hat{\beta}_2(x_{i2} - \hat{\mu}_2) + \dots \quad (5)$$

The  $\hat{y}_i$  term is nothing more than a composite of all the student's past scores. It is a one-number summary of the student's level of achievement prior to the current year, and this is sometimes referred to as the expected score, entering score, or predicted achievement. The different prior test scores making up this composite are given different weights (by the regression coefficients, the  $\hat{\beta}$  terms) to maximize its correlation with the response variable. Thus, a different composite would be used when the response variable is Math than when it is Reading, for example.  $\hat{y}_i$  represents prior achievement before the effect of the current district, school, or teacher.

For the purposes of the TMA program, the  $\hat{y}_i$  terms, expressed in scale scores, are compared to the observed scale score for each student for each assessment to determine whether they have exceeded their predicted achievement. These are then aggregated to the teacher level for each applicable assessment for teachers using teacher-student linkages like those from the R Verification process. These aggregations are produced for each assessment for each teacher, and all students across all grades and subjects for which the teacher has results are aggregated into a single composite. The composite value is used for the purposes of TMA. This percentage is calculated as follows:

$$\%MEE = \frac{\sum_{i=1}^n (w_i S_i)}{\sum_{i=1}^n w_i}, \text{ where } S_i = \begin{cases} \text{if } y_i > \hat{y}_i \\ \text{if } y_i \leq \hat{y}_i \end{cases} \quad (6)$$

When calculating this percentage, a teacher's percentage of instructional responsibility, or  $w_i$ , for student  $i$  is used to weight the results. If different students are claimed at different percentages of

---

instructional responsibility for a given teacher, each is weighted proportionally when calculating the percentage of students exceeding predicted achievement for that teacher. In equation 6,  $S_i = 1$  if the observed scale score ( $y_i$ ) for that student is greater than the predicted achievement for that student, and  $S_i = 0$  if the observed scale score ( $y_i$ ) for that student is less than or equal to the predicted achievement for that student. These values are then multiplied by the instructional responsibility percentage in the numerator, and that student's contribution to the denominator is the instructional responsibility percentage. This is done for all students for a given assessment and for the composite that goes across all assessments for which the teacher has results. The resulting numerator and denominator include all applicable students to calculate a percentage of students meeting or exceeding expectations.

For example, if a teacher has 30 students that all exceeded their predicted achievement and were all claimed at 100 percent instructional responsibility, each student contributes values of 1 to both the numerator and denominator, resulting in 30 divided by 30, or 100% of students exceeding predicted achievement. If the same teacher had 50 percent instructional responsibility for the same set of students, the numerator would be 15 and the denominator would be 15, also resulting in 100% of students exceeding predicted achievement.

Alternatively, in some cases some students for a teacher might be claimed at 50 percent instructional responsibility and others might be claimed at 100 percent instructional responsibility. The students with 50% instructional responsibility for the teacher contribute 0.5 to the numerator and denominator if they exceed predicted achievement or 0 to the numerator and 0.5 to the denominator if their observed scale score is less than or equal to predicted achievement. The students with 100% instructional responsibility for the teacher contribute 1 to the numerator and denominator if they exceed predicted achievement or 0 to the numerator and 1 to the denominator if their observed scale score is less than or equal to predicted achievement. In practice, this means that students that have higher percentages of instructional responsibility have a proportionally larger contribution to a teacher's percentage of students exceeding expectations than those that have lower percentages of instructional responsibility.