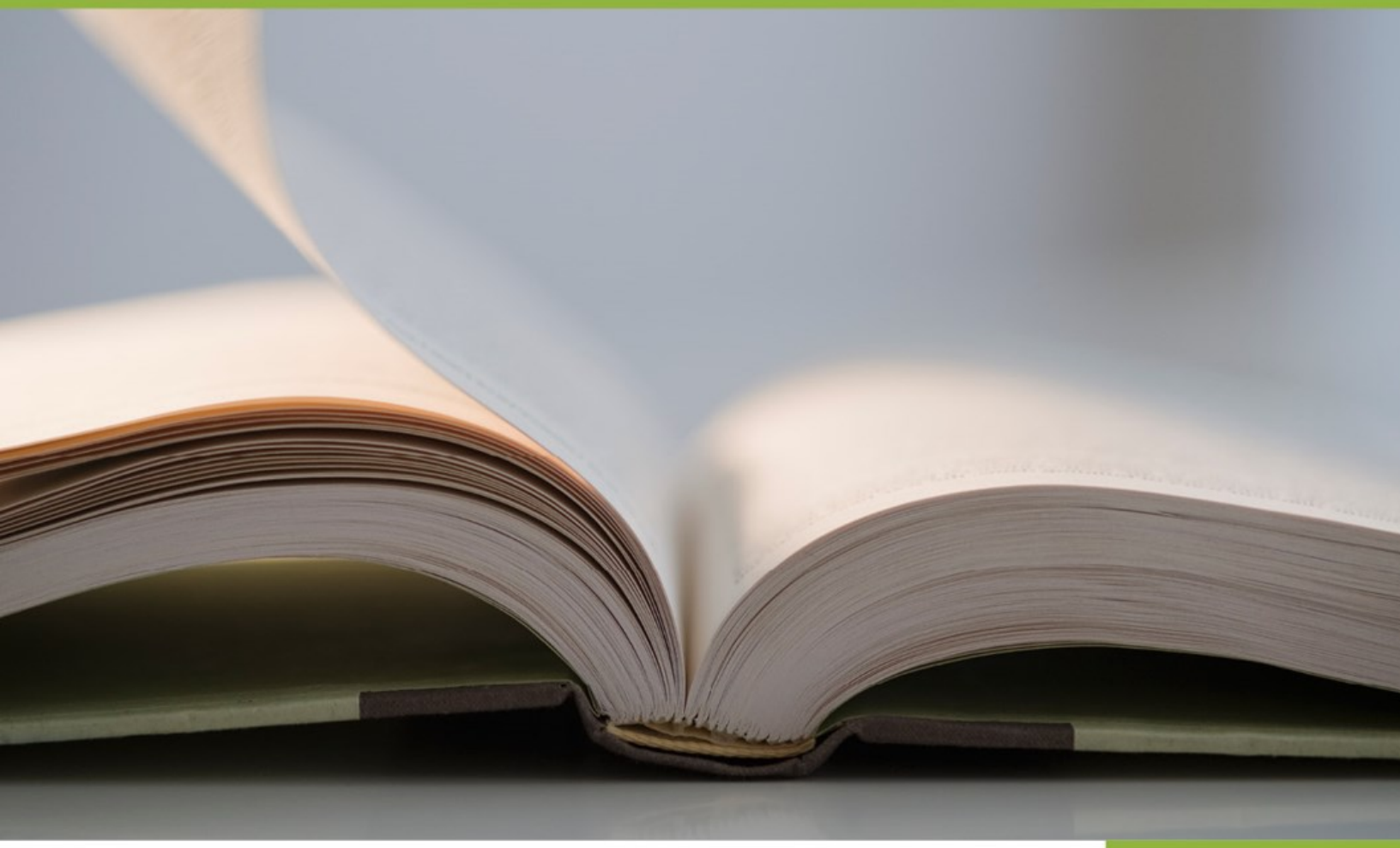
SAS® EVAAS

Statistical Models and Business Rules

Prepared for Texas Tech University and the Texas Education Agency



sas
THE POWER TO KNOW.

Contents

1 Introduction

1.1 Background and Purpose

In June 2019, the Texas Legislature passed House Bill 3, which established the Teacher Incentive Allotment (TIA). This new initiative aims, in part, to recruit and retain excellent teachers, and for participating Local Education Agencies (LEAs) to develop their own designation system in support of these goals. The legislation requires that LEAs submit their systems to the Texas Education Agency (TEA) for review and approval of the required components.

This document focuses on one required component of the local designation system: student growth measures. As part of the system review by Texas Tech University (TTU) and TEA, local student growth models can be compared to statewide models to verify their validity and relative accuracy. This document outlines the statewide statistical growth models that will be used as part of the comparison: two predictive value-added models. The document includes a technical description of the models, an explanation of expected growth within the models, and how model outputs can be used to classify teachers in accordance with TIA. These sections are followed by details outlining the data used and business rules.

The goal of this document is to provide clarity into the statewide student growth models that are compared to data submitted from local designation systems.

2 Predictive Model

2.1 Overview

The predictive model is a regression-based value-added model where growth is a function of the difference between students' expected scores and their actual scores. Expected growth is met when students with a district, school, or teacher made the same amount of growth as with the average district, school, or teacher.

In more technical terms, the predictive model used here is sometimes known as the univariate response model (URM), a linear mixed model, and, more specifically, an analysis of covariance (ANCOVA) model.

Conceptually, growth in the predictive model is simply the difference between students' entering and exiting achievement. If students score where they were expected to score, then the growth measure will be zero (or close to zero). Zero represents "expected growth." Positive growth measures are evidence that students made more than the expected growth, and negative growth measures are evidence that students made less than the expected growth.

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 expected growth in advance of the assessment being taken by students. The predictive model defines expected growth within each year.

More specifically, expected growth means that a teacher's students made the same amount of growth as students with the average teacher in the state for that same year, subject, and grade when considering students' prior testing history and additional student-level and group-level factors. Growth measures tend to be centered on expected growth every year with approximately half of the teacher estimates above zero and approximately half of teacher estimates below zero.

2.2 Technical Description

In the predictive model, each student receives an expected score based on their own prior testing history. In practical terms, the expected score represents the student's entering achievement because it is based on all prior testing information to date.

The expected scores can be aggregated to a specific teacher and then compared to the students' actual scores. In other words, the growth measure is a function of the difference between the average exiting score (or actual scores) and the average entering score (or expected score) for a group of students. The expected scores are reported in the scaling units of the test.

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.
- The categorical variables (α terms, class variable, factor) are the teachers from whom the student received instruction in the subject, grade, and year of the response variable (y).

Algebraically, the model can be represented as follows for the i^{th} student when there is no team teaching.

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

In the case of team teaching, the single α_j is replaced by multiple α terms, each multiplied by an appropriate weight. The μ terms are means for the response and the predictor variables. α_j is the teacher effect for the j^{th} teacher—the teacher who claimed responsibility for the i^{th} student. The β terms are regression coefficients. Predictions to the response variable are made by using this equation with estimates for the unknown parameters (μ terms, β terms, and sometimes α_j). 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:

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

Two difficulties must be addressed to implement the predictive model. First, not all students will have the same set of predictor variables due to missing test scores. Second, the estimated parameters are pooled within teacher. 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 projection 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). These are not simply the grand mean scores. It can be shown that in an ANCOVA if we impose the restriction that the estimated teacher effects should sum to zero (that is, the teacher effect for the "average teacher" is zero), then the appropriate means are the means of the teacher means. The teacher means are obtained from the EM algorithm mentioned above, which accounts for missing data. The overall means ($\hat{\mu}$ terms) are then obtained as the simple average of the teacher means.

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 has a minimum of three prior test scores.

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

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 or entering score. 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. Note that the $\hat{\alpha}_j$ term is not included in the equation. Again, this is because $\hat{y}_i$ represents prior achievement before the effect of the current district, school, or teacher.

The second step in the predictive model is to estimate the teacher effects (α_j) using an ANCOVA model as shown in Equation 6.

$$y_i = \gamma_0 + \gamma_1 \hat{y}_i + \gamma_2 + \gamma_3 x_s + \sum_{t=1}^{5} \gamma_{4.t} w_{i.t} + \gamma_5 \overline{w}_s + \alpha_j + \epsilon_i \quad (6)$$

Because the model adjusts for additional student and group-level characteristics, it includes terms for students' census tier, school rurality, average school census tier, and the teacher-level average expected score. $\overline{\hat{y}}_j$ is the teacher-level mean of $\hat{y}_i$, which represents average entering achievement. x_s represents school-level rurality, with a value of 1 for schools identified as rural and 0 for schools not identified as rural. $w_{i.t}$ is an indicator of student census tier with *t* taking on values of 1, 2, 3, 4, or 5 with *t*=0 as a reference category. $\overline{w}_s$ is the school mean of the student-level census tier values.

Note that prior test scores used in these models do not need to be on the same scale as the assessment being predicted. Just as height (reported in inches) and weight (reported in pounds) can predict a child's age (reported in years), the predictive model can use test scores from different scales to find the predictive relationship.

2.3 Model Outputs

2.3.1 Grades and Subjects

Based on the data received and described in Section 3.1, the predictive model provides student growth measures for teachers in the following assessed areas:

- Mathematics, grades 4–8
- Reading Language Arts (RLA), grades 4–8
- Science, grades 5 and 8
- Social Studies, grade 8
- Algebra I
- Biology
- English I
- English II
- US History

These measures can, in turn, be used and interpreted in different ways to assess the significance of growth made by students taught by a specific teacher. The output used to support TIA is the percentage of students meeting or exceeding expectations, which is described in more detail in Section 2.3.2. In addition to providing this metric in each individual subject and grade or course, an overall measure is created that spans across all subjects, grades, and course taught by a teacher each year.

2.3.2 Percentage of Students Meeting or Exceeding Expectations

As described in Section 2.2, the predictive model produces an expected scale score ($\hat{y}$) for each student included in the model. For the purposes of TIA, all available expected student scale scores a given school year are compared to students' actual scale scores to determine which students met or exceeded the

expected scale score. These are then aggregated to the teacher level across all available grades and subjects for the teacher to generate a single value using the following equation:

$$\frac{\text{Number of student scale scores greater than or equal to expected scale scores}}{\text{Number of student scale scores}} \quad (8)$$

For example, if a teacher had 60 student scale scores included in the model across grades and subjects and 48 met or exceeded the expected scale score, then the calculation of this metric would be:

$$\frac{48}{60} = .80 = 80\% \text{ of students met or exceeded expectations} \quad (9)$$

To create an overall measure, all students are used in each subject and grade or course connected to a teacher, and an overall percentage of students that have scored greater than or equal to their expected score is calculated.

The overall measure including all available subjects and grades or courses is used to support the data validation checks performed by TTU on data submitted by districts. For use in the applicable data validation checks, the overall measures for teachers are assigned an overall category using performance standards determined by TEA. Teachers with percentages below 55 are categorized as “Not Designated,” teachers with percentages of 55 or greater and less than 60 are categorized as “Recognized,” teachers with percentages of 60 or greater and less than 70 are categorized as “Exemplary,” and teachers with percentages of 70 or greater are categorized as “Master” for the purposes of the relevant data validation checks.

3 Data Received and Data Processing Business Rules

3.1 Data Received

TEA provides STAAR EOG Reading (through 2021-22)/RLA (2022-23 forward) and Math data for grades 3–8, STAAR EOG Science data for grades 5 and 8, STAAR EOG Social Studies data for grade 8, STAAR Writing data for grades 4 and 7 (through 2021-22), and EOC assessment data (English I/II, Algebra I, Biology, US History) from the 2012-13 school year to present. TEA also provides teacher-student linkages for the purpose of connecting students to teachers in the modeling.

In addition to assessment score results and student teacher linkages, TEA provides additional student and campus level characteristics for use in the adjusted model. These include:

- Student Economic Disadvantaged Code
- Student Census Tier
- Campus Rurality Indicator

3.2 Entity Resolution

SAS connects students across multiple years of data received from TEA using student identification variables. These variables are Last Name, First Name, Birth Date, Unique ID, and Local Student ID.

3.3 Data Processing Business Rules

3.3.1 Course to Assessment Mapping of Linkages

Teacher-student linkages are connected to specific assessments based on a course to subject mapping approved by TEA.

3.3.2 Dropping Unused Linkages

Teacher-student linkages that are not successfully mapped to an assessed subject are not retained.

3.3.3 Exclusion of STAAR Version T Records

STAAR version T assessment records are excluded.

3.3.4 Exclusion of Non-Scorable Assessment Records

Non-scorable assessment results are excluded.

3.3.5 Exclusion of Retest Assessment Records

EOC retest assessments records are excluded. More specifically, records marked as retests are removed, and then any remaining records that are not the first record for that student for that EOC subject are also removed. For any student with multiple test records on a STAAR grade level assessment within a school year, only the record with the earliest test date is used.

3.3.6 Exclusion of June and July Records

The small number of records from assessments administered in June and July are not currently included in the data provided to SAS. As a result, these records are excluded from the analysis.

3.3.7 Exclusion of Raw Scores of Zero

Records with raw scores of zero are excluded.

3.3.8 Adjustment of Grade 3–5 Spanish STAAR RLA and Mathematics Records

Spanish assessment scores might be adjusted using Deming regression such that the gains of students transitioning from Spanish-to-English are equivalent to students transitioning from English-to-English. This adjustment is applied for each combination of subject, grade, year, test language, and scale score if needed.

3.3.9 Minimum Number of Prior Assessment Scores

For most grades or subjects, three prior assessment scores are required for a student to be included in the predictive model. The only exceptions are assessments in grade 4, which require only two prior assessment scores. Note that the required scores do not necessarily need to include a score from the prior year in the same subject area, as the model can use the available prior scores and accommodate missing data.

3.3.10 Outlier Detection

Student assessment scores are checked to determine whether they are outliers in context with all other scores in a reference group of scores from the individual student. These reference scores are weighted differently depending on proximity in time to the score in question. Scores are checked for outliers using related subjects as the reference group. For example, when searching for outliers for Math test scores, all Math subjects are examined simultaneously. Any scores that appear inconsistent, given the other scores for the student, are flagged. Scores are flagged in a conservative way to avoid excluding any student scores that should not be excluded. Scores can be flagged as either high or low outliers. Once an outlier is discovered that outlier will not be used in the analysis.

This process is part of a data quality procedure to ensure that no scores are used if they were in fact errors in the data, and the approach for flagging a student score as an outlier is fairly conservative.

Considerations included in outlier detection are:

- Is the score in the tails of the distribution of scores? Is the score very high or low achieving?
- Is the score “significantly different” from the other scores, as indicated by a statistical analysis that compares each score to the other scores?
- Is the score also “practically different” from the other scores? Statistical significance can sometimes be associated with numerical differences that are too small to be meaningful.
- Are there enough scores to make a meaningful decision?

To decide whether student scores are considered outliers, all student scores are first converted into a standardized normal Z-score. Then each individual score is compared to the weighted combination of all the reference scores described above. The difference between these two scores provides a t-value of each comparison. Using this t-value, SAS can flag individual scores as outliers.

There are different business rules for the low outliers and the high outliers, and this approach is more conservative when identifying high outliers.

For low-end outliers, the rules are:

- The percentile of the score must be below 50.
- The t-value must be below -3.5 when looking at the difference between the score in question and the reference group of scores within the same subject and/or below -4.0 when comparing to the reference group of scores across all subjects.

- The percentile of the comparison score must be above a certain value. This value depends on the position of the individual score in question but will range from 10 to 90 with the ranges of the individual percentile score.

For high-end outliers, the rules are:

- The percentile of the score must be above 50.
- The t-value must be above 4.0 when comparing to the reference group of scores within the same subject and/or above 5.0 when comparing to the reference group of scores across all subjects.
- The percentile of the comparison score must be below a certain value.
- There must be at least three scores in the comparison score average.

3.3.11 Minimum Number of Students for Teacher Growth Data

To generate a teacher growth measure for the predictive model in a given grade/subject/year, the teacher must have at least five full-time equivalent (FTE) students included in the model. The teacher's number of FTE students is based on the number of students linked to that teacher and the percentage of instructional time the teacher has for each student. For example, if a teacher taught 10 students for 50% of their instructional time, then the teacher's FTE number of students would be five, and they would meet the minimum for receiving a teacher growth measure.

3.3.12 Student Census Tier

The student-level census tier that is used as a covariate in the model is a 1 to 5 socio-economic indicator created for students identified as economically disadvantaged. Tier 1 is the highest socio-economic group, and Tier 5 is the lowest socio-economic group. Any student whose Economic Disadvantaged Code is zero (not identified as economically disadvantaged) has their Student Census Tier set to zero. Lastly, when tier information for a student is not provided, the value for Student Census Tier is set to zero.

3.3.13 Average School Census Tier

The average school census tier used as a covariate in the model is calculated for a given school/subject/grade using students that are included in the teacher value-added model.

3.3.14 Campus Rurality Indicator

The campus rurality indicator that is used as a covariate in the model identifies if a campus is categorized as rural and can be a value of "Yes" or "No". When the campus rurality information for a campus is not provided, the rurality value for that campus is set to "No".

3.3.15 Teacher-Level Average Expected Score

The teacher-level average expected score that is used as a covariate in the model is calculated only for teachers with at least five FTE students and is weighted by instructional responsibility.