Contents

1 Introduction ....................................................................................................................... 2
  1.1 Background and Purpose ......................................................................................... 2
  1.2 Measuring Student Growth .................................................................................... 2

2 Predictive Model ............................................................................................................... 4
  2.1 Overview .................................................................................................................. 4
  2.2 Technical Description ............................................................................................ 4
  2.3 Model Outputs ......................................................................................................... 6
    2.3.1 Grades and Subjects ......................................................................................... 6
    2.3.2 Percentage of Students Meeting or Exceeding Expectations ......................... 7
    2.3.3 Effect Size ....................................................................................................... 7
    2.3.4 Test Statistic .................................................................................................... 8

3 Student Growth Percentiles (SGPs) ............................................................................. 9
  3.1 Overview .................................................................................................................. 9
  3.2 Technical Description ............................................................................................ 9
  3.3 Model Outputs ......................................................................................................... 12
    3.3.1 Grades and Subjects ......................................................................................... 12
    3.3.2 Description of Model Output ........................................................................... 12

4 Data Received and Data Processing Business Rules ................................................... 13
  4.1 Data Received .......................................................................................................... 13
  4.2 Entity Resolution ..................................................................................................... 13
  4.3 Data Processing Business Rules ............................................................................ 13
    4.3.1 Course to Assessment Mapping of Linkages ...................................................... 13
    4.3.2 Dropping Unused Linkages .............................................................................. 13
    4.3.3 Exclusion of STAAR Version T Records ......................................................... 13
    4.3.4 Exclusion of Non-Scorable Assessment Records .............................................. 13
    4.3.5 Exclusion of Retest Assessment Records ........................................................ 13
    4.3.6 Exclusion of Raw Scores of 0 ......................................................................... 13
    4.3.7 Adjustment of Grade 3-5 Spanish STAAR Reading and Mathematics Records .. 13
    4.3.8 Minimum Number of Prior Assessment Scores .............................................. 14
    4.3.9 Outlier Detection .............................................................................................. 14
    4.3.10 Minimum Number of Students for Teacher Growth Data ................ ............... 15
    4.3.11 SGPs and Additional Requirement for Students to be Included ....................... 15
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 participating Local Education Agencies (LEAs) develop their own designation system to identify these teachers to receive an allotment. 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 TEA, local student growth models can be compared to statewide models to verify their validity and relative accuracy. This document outlines two statewide statistical growth models that will be used as part of the comparison: a predictive value-added model and Student Growth Percentiles (SGPs). The document includes a technical description of each model, an explanation of expected growth within each model, 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 and their use in comparisons to local systems.

1.2 Measuring Student Growth

There are many different ways of measuring student growth. One useful way to describe them at a high level is to define what is meant by growth and how an expectation of growth is defined. From here, there are other specifics about how each methodology is implemented and what statistical procedures are used.

In general, the approaches used in this document describe growth by looking at entering achievement as defined by previous test scores and then compare that to current achievement. This difference is looked at in relation to the statewide average change. In other words, the expectation is based on a statewide average in each methodology, and growth is a comparison of entering and exiting achievement. The way growth is described is unique quantitatively within each methodology and will be described in each section.

Although their approaches differ, both the predictive model and SGPs are regression models. Regression models are used in virtually every field of study, and their intent is to identify relationships between two or more variables. When it comes to measuring growth, regression models identify the relationship between prior test performance and actual test performance for a given course. These models have been reviewed and replicated by a variety of independent value-added experts,¹ and they have been

¹ See, for example, Katherine E. Castellano and Andrew D. Ho, A Practitioner’s Guide to Growth Models (Council of Chief State School Officers, 2013) or the independent review by WestEd in North Carolina’s approved Consolidated State Plan under the Every Student Succeeds Act (p. 39-40).
used for teacher evaluation, school accountability, and merit-pay systems in many states over the past decade or more.
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 with their actual scores. Expected growth is met when students with a district, school, or teacher made the same amount of growth as students in 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. 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 his or her 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 \(x_i\), predictor variables, explanatory variables, independent variables) are scores on tests the student has already taken.
- The categorical variable \(\alpha_i\), 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.
\[ y_i = \mu_y + \alpha_j + \beta_1(x_{i1} - \mu_1) + \beta_2(x_{i2} - \mu_2) + \cdots + \epsilon_i \]  \hfill (1)

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

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

Two difficulties must be addressed in order 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,

\[ C = \begin{bmatrix} C_{yy} & C_{yx} \\ C_{xy} & C_{xx} \end{bmatrix} \]  \hfill (3)

This matrix is estimated using the Expectation Maximization 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:

\[ \hat{\beta} = C_{xx}^{-1}C_{xy} \]  \hfill (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 that 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.
\[ \hat{y}_i = \hat{\mu}_y + \hat{\beta}_1(x_{i1} - \hat{\mu}_1) + \hat{\beta}_2(x_{i2} - \hat{\mu}_2) + \cdots \]  

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 term is called the expected score or entering score in the web reporting. The different prior test scores making up this composite are given different weights (by the regression coefficients, the \( \hat{\beta}_s \)) in order 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 \( \hat{a}_j \) 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 (\( \alpha_j \)) using the following ANCOVA model.

\[ y_i = \gamma_0 + \gamma_1 \hat{y}_i + \alpha_j + \epsilon_i \]  

In the predictive model, the effects (\( \alpha_j \)) are considered random effects. Consequently, the \( \hat{\alpha}_j \)s are obtained by shrinkage estimation (empirical Bayes).\(^2\) The regression coefficients for the ANCOVA model are given by the \( \gamma \)s.

Note that prior test scores 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 4.1, the predictive model provides student growth measures for teachers in the following assessed areas:

- Mathematics, grades 4-8
- Reading, grades 4-8
- Science, grades 5 and 8
- Social Studies, grade 8
- Writing, grades 4 and 7
- 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 TIA system review includes three different ways that are outlined below and then described in more detail:

- Percentage of students meeting or exceeding expectations
- Effect size
- Test statistic

In addition to providing these metrics in each individual subject/grade or course, an overall measure is sometimes 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 ($\tilde{y}$) for each student included in the model. For the purposes of TIA, all available expected student scale scores from the 2018-19 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:

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

(7)

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, the calculation of this metric would be:

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

(8)

To create an overall measure, all students are used in each subject and grades 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.

### 2.3.3 Effect Size

The student standard deviation of growth can be used to provide context about the magnitude of growth being made by a group of students. For the predictive model, students have an actual score and an expected score. The difference is a measure of student growth. The standard deviation of the student-level distribution of growth is available for each year, subject, and grade. Dividing the teacher growth measures by the student standard deviation provides a value known as an “effect size,” and it indicates the practical significance regarding the group of students and whether they met, exceeded, or fell short of expected growth. The effect size is calculated as:

$$\frac{\text{Growth Measure}}{\text{Standard Deviation (of student-level distribution of growth)}}$$

(9)

The overall effect size is calculated by averaging the individual effect sizes in each subject/grade and course with weighting defined by the effective number of students in each of those measures.
2.3.4 Test Statistic

The standard error can help indicate whether a value-added estimate is significantly different from expected growth. This determining value is a test statistic and is typically referred to as the growth index. The calculation is simply the value-added measure divided by its standard error. Since the expectation of growth is zero, this measures the certainty about the difference of a growth measure and zero. Stated differently, it conveys the level of evidence that a teacher’s students, on average, exceeded or fell short of the expected growth.

These test statistics can be described as confidence intervals around value-added estimates. For example, if the test statistic is about two, the 95 percent confidence interval does not cross zero, our expectation of growth. This indicates that based on the available evidence, there is a 95% chance that the teacher’s students, on average, exceeded expected growth.

The test statistic can be expressed as:

\[
\text{Index or Test Statistic} = \frac{\text{Growth Measure}}{\text{Standard Error}}
\]

(10)

There are a few methods of creating an index value that goes across all subjects and grades, and we describe one such method. Although a test statistic is simply the growth measure divided by the standard error, it is equivalent to the effect size divided by the standard error of the effect size. Since growth measures are on different scales, the effect sizes can be averaged to get an overall effect size due to them being on the same scale. A standard error can be calculated for this averaged effect size and a new test statistic of this overall value can be generated by doing the division.
3 Student Growth Percentiles (SGPs)

3.1 Overview
Student Growth Percentiles (SGPs)\(^3\) are another approach to measuring the growth of a teacher’s students. SGPs compare the growth of students to other students who have similar prior academic achievement. Results from the SGP model are typically reported based on the median SGP across a teacher’s students. To obtain the median, SGPs are ordered from least to greatest for a given teacher, and the middle, referred to as the median, value is selected. This section explains the technical details of the SGP model and draws some distinctions between this approach and the predictive model.

As with the predictive model, the SGP model compares students’ actual scale scores to an expectation. In both models that expectation comes from a multiple regression equation but using different methodologies as explained in the next section. Conceptually, the expectation allows each student to be compared to other students who have the same prior assessment scores. The expectation corresponds to the “average” score for students having the same prior assessment scores. (In neither model is an individual student’s actual scale score compared to the actual scale scores of all students with the same prior achievement despite what some descriptions of the SGP model might say.) The models differ in how they express the comparison between the actual score and the expectation. The predictive model takes the difference between actual score and expectation so that the growth expectation is zero. In the SGP model, the comparison is expressed as a percentile so that the growth expectation is 50. In both models, when the actual score is equal to the expectation, the interpretation is that a student’s actual scale score was higher than it was expected to be based on the statewide relationships of that score and prior tests.

3.2 Technical Description
The SGP model is conceptually similar to the predictive model. It is a regression-based model that measures student growth in terms of the difference between students’ expected scores and their actual scores in a particular grade, subject, and year.

The predictive model equation in Section 2.2 can be rewritten in the form of a multiple regression equation by gathering all the \(\hat{\mu}\) terms together into a single intercept term:

\[
\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots
\]  

(11)

This student’s growth is represented by \((y_i - \hat{y}_i)\). In a regression analysis, this is called a residual. This growth is expressed in scale score units. The effects in the predictive model are also in scale score units because they are calculated by averaging the residuals (and applying shrinkage estimation to the result).

An alternative way to express a student’s growth is in percentile units. Among students having the same prior test scores (that is, the same values for \(x_{1}, x_{2}, \text{etc.}\)) and therefore having the same predicted score

\[^{\text{For more information about SGPs, please see Betebenner, Damian W., A Technical Overview of the Student Growth Percentile Methodology: Student Growth Percentiles and Percentile Growth Projections/Trajectories (Dover, N.H.: National Center for the Improvement of Educational Assessment, 2011).}}\]
\( \hat{y} \), what percentage of that group of students scored lower than this specific student? The result is an SGP. A predictive model version of the SGP can be calculated as follows.

\[
SGP_{URM} = 100 \Phi[(y_i - \hat{y}_i)/SE(\hat{y}_i)]
\]

(12)

This calculation assumes that scores are normally distributed since the \( \Phi \) in the formula is the standard normal distribution function.

The SGP model, as usually implemented, differs from the predictive model in several important respects that involve relaxing certain assumptions that are made in regression modeling (including the predictive model). First, the normality assumption is avoided by using quantile regression. To obtain SGPs, a separate prediction equation, with its associated \( \hat{y}_i \) value for each student, is produced for each of the percentiles 1 through 99. The SGP for the \( i \)-th student is defined as the percentile associated with the equation whose \( \hat{y}_i \) is closest to \( y_i \).

Quantile regression also avoids another of the usual assumptions made in regression modeling: the constant variance assumption (homoscedasticity). The assumption is that the amount of variation in scores in the population of students having identical prior test scores is the same for every such population.

Another assumption of regression modeling is linearity. In the context of multiple regression, what this actually means is that the prediction equation, such as the one given above, can be written as a sum of terms with each term consisting of a known value (\( x_{i1} \) for example) multiplied by a coefficient that has to be estimated (\( \hat{\beta}_1 \)). Nonlinear relationships are easily accommodated by incorporating polynomial terms such as \( x_i^2 \) and cross-product terms such as \( x_1 x_2 \). It has not been found necessary to include such terms in URM models, but they are routinely used in the standard implementation of the SGP model. Specifically, each prior test score is expanded from a single term to seven terms known as a cubic spline with four knots. The specific terms used in the SGP model are called B-splines, but conceptually the expansion of \( \hat{\beta}_1 x_{i1} \) is equivalent to the following:

\[
\begin{align*}
\hat{\beta}_{1.1} x_{i1} + \hat{\beta}_{1.2} x_{i1}^2 + \hat{\beta}_{1.3} x_{i1}^3 + \hat{\beta}_{1.4} (x_{i1} - k_1)^2 + \hat{\beta}_{1.5} (x_{i1} - k_2)^2 + \hat{\beta}_{1.6} (x_{i1} - k_3)^2 + \hat{\beta}_{1.7} (x_{i1} - k_4)^2
\end{align*}
\]

(13)

\( k_1 \) through \( k_4 \) are the knots that are assigned the values of the 20th, 40th, 60th, and 80th percentiles of \( x_{i1} \), respectively. The expression \((x - k)_+\) has the value of \((x - k)\) if it is greater than zero; otherwise it is set to zero.

Finally, it should be noted that the criterion used in multiple linear regression is least squares. That is, the coefficients (the \( \hat{\beta} \) values) are chosen to minimize the sum of the squared residuals, \( \Sigma(y_i - \hat{y}_i)^2 \). Quantile regression uses a different criterion, an extension of the sum of the absolute values of the residuals, \( \Sigma|y_i - \hat{y}_i| \), which is modified to accommodate different equations at different percentiles. Specifically, the coefficients for the 100\( p \)-th percentile equation minimize

\[
\Sigma p|y_i - \hat{y}_i|_+ + (1 - p)|y_i - \hat{y}_i|_-
\]

(14)

where \( |y_i - \hat{y}_i|_+ = (y_i - \hat{y}_i) \) if greater than zero, otherwise it is zero, and \( |y_i - \hat{y}_i|_- = -(y_i - \hat{y}_i) \) if
\((y_i - \hat{y}_i)\) is less than zero, otherwise it is zero. When \(p = 0.5\) (so-called median regression), this simplifies to \(\sum|y_i - \hat{y}_i|\). Using absolute values of residuals makes quantile regression more robust against extreme values (outliers) in the response \(y\). However, it does not protect against outliers in the predictors (\(x\) values).

In the discussion of the predictive model, it was noted that there are two difficulties that must be addressed. First, not every student has the same set of prior test scores. Second, the predictive model is an analysis of covariance model, which means that the grouping of students within districts or schools or teachers is taken into account during modeling. The SGP model must also address these difficulties, but it does so differently than the predictive model.

The SGP model handles missing predictors as follows. For a particular response variable (e.g., eighth-grade Math), the SGP model calculates regression parameters for each different set of predictor variables (previous test scores). To keep this manageable, the number of regressions is limited. First, the predictor variables must be in the same subject as the response. For example, only prior Math scores are used when the response is Math. Second, an “older” predictor is used only when all of the more recent predictors are also available. For example, with Math in grade 8 as the response, an SGP model would be run with each of the following sets of predictors: \{math7\}, \{math7, math6\}, \{math7, math6, math5\}. In each case, all available students are used, which means all who have an eighth-grade Math score in the current year along with a value for predictor. Consequently, the students used in a given model are a subset of the students used in models having fewer predictors. For example, the \{math7\} model would include all students having an earlier seventh-grade Math score, the \{math7, math6\} model would include all students having both an earlier seventh-grade Math score and an earlier sixth-grade Math score, etc. The SGP for a particular student is taken from the model that has the most previous test scores.

The second difficulty is not directly addressed by SGP models. During modeling to obtain the student level SGPs, the grouping of students within districts, schools, or teachers is ignored. It is only after modeling and after an SGP is obtained for each student, that a growth measure is obtained for each group (district, school, teacher). The growth measure is the median of the SGPs for students in the group. For teachers, if there is co-teaching with each teacher claiming some fractional responsibility for each student, a weighted median is calculated using the fractional responsibilities as weights.

There is an additional difficulty that is particular to the SGP model. A mentioned above, a quantile regression model is run for each of the 99 percentiles from 1 to 99. The SGP for a particular student is defined to be the percentile associated with the model whose \(\hat{y}_i\) is closest to \(y_i\). However, the computational algorithm used in quantile regression is such that this can occur for multiple percentiles. The SGP for the \(i\)-th student is therefore not uniquely defined. In such a case, the median of the percentiles is assigned as the SGP unless the percentiles have a range of greater than seven, in which case no SGP is assigned. For example, if a student’s best-fitting equations occurred at percentiles 41, 42, 43, 44, and 45, the SGP would be 43. But if the best-fitting equations occurred at percentiles 40, 41, 42, 43, 44, 45, 46, 47, and 48, no SGP would be assigned.
### 3.3 Model Outputs

#### 3.3.1 Grades and Subjects
Based on the data received and described in Section 4.1, SGPs provide student growth measures for teachers in the following assessed areas:

- Mathematics, grades 4-8
- Reading, grades 4-8
- Science, grades 8
- Writing, grade 7
- Algebra I
- Biology
- English I
- English II
- US History

In contrast to the predictive model, SGP data is not calculated for grade 4 Writing, grade 5 Science, or grade 8 Social Studies because there is not a prior assessment in those subject areas for use in the SGP model.

#### 3.3.2 Description of Model Output
As described above, each student is assigned an SGP. For the purposes of analysis to support TIA, the resulting student-level SGPs are aggregated to the teacher level across all available grades and subjects. Specifically, the median value across all students in all grades and subjects is identified for each teacher. If there is co-teaching with each teacher claiming some fractional responsibility for each student, a weighted median is calculated using the fractional responsibilities as weights. If $x_i$, $i = 1, ..., n$ are the SGPs for a teacher’s students, then

$$\text{Median}(x) = \begin{cases} 
0.5 (x_i + x_{i+1}) & \text{if } \sum_{j=1}^{i} w_j = (W/2) \\
x_{i+1} & \text{if } \sum_{j=1}^{i} w_j < (W/2) < \sum_{j=1}^{i+1} w_j 
\end{cases} \quad (15)$$

where $w_j$ is the weight (fractional responsibility) associated with $x_j$ and $W = \sum_{i=1}^{n} w_i$ is the sum of the weights.
4 Data Received and Data Processing Business Rules

4.1 Data Received
TEA provided STAAR EOG Reading and Math data for grades 3–8, STAAR EOG Science data (grades 5 and 8), STAAR EOG Social Studies data (grade 8), STAAR Writing data (grades 4 and 7), and EOC assessment data (English I/II/III, Algebra I/II, Biology, US History) from the 2014-15, 2015-16, 2016-17, 2017-18, and 2018-19 school years. The only exception is that English III and Algebra II data was not available for 2014-15. TEA also provided teacher-student linkages for the 2016-17, 2017-18, and 2018-19 school years.

4.2 Entity Resolution
SAS connected students across the five years of data received from TEA using student characteristic variables. These variables were last name, first name, birth date, Unique ID, and SSN.

4.3 Data Processing Business Rules

4.3.1 Course to Assessment Mapping of Linkages
Teacher-student linkages were connected to specific assessments based on a course to subject mapping approved by TEA.

4.3.2 Dropping Unused Linkages
Teacher-student linkages that are not successfully mapped to an assessed subject are not retained.

4.3.3 Exclusion of STAAR Version T Records
STAAR version T assessment records are excluded.

4.3.4 Exclusion of Non-Scorable Assessment Records
Non-scorable assessment results are excluded.

4.3.5 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 were not the first record for that student for that EOC subject are also removed. For any duplicate test records in STAAR grade level assessments, only the record with the earliest test date was used.

4.3.6 Exclusion of Raw Scores of 0
Records with raw scores of 0 are excluded.

4.3.7 Adjustment of Grade 3-5 Spanish STAAR Reading and Mathematics Records
Spanish assessment scores are 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.
4.3.8 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.

The SGP model uses one, two, or three prior scores within the same subject area. As a result, the SGP model requires a minimum of one prior score in the same subject area.

4.3.9 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 of 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

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.

### 4.3.10 Minimum Number of Students for Teacher Growth Data

In order to generate a teacher growth measure for the predictive model or SGPs 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.

### 4.3.11 SGPs and Additional Requirement for Students to be Included

In some cases, SGPs might not be assigned to specific students if certain conditions are met. The SGP model uses a quantile regression model is run for each of the 99 percentiles from 1 to 99. The SGP for a particular student is defined to be the percentile associated with the model whose $\hat{y}_i$ is closest to $y_i$. However, the computational algorithm used in quantile regression is such that this can occur for multiple percentiles. The SGP for the $i$-th student is therefore not uniquely defined. In such a case, the median of the percentiles is assigned as the SGP unless the percentiles have a range of greater than seven, in which case no SGP is assigned. For example, if a student’s best-fitting equations occurred at percentiles 41, 42, 43, 44, and 45, the SGP would be 43. But if the best-fitting equations occurred at percentiles 40, 41, 42, 43, 44, 45, 46, 47, and 48, no SGP would be assigned.