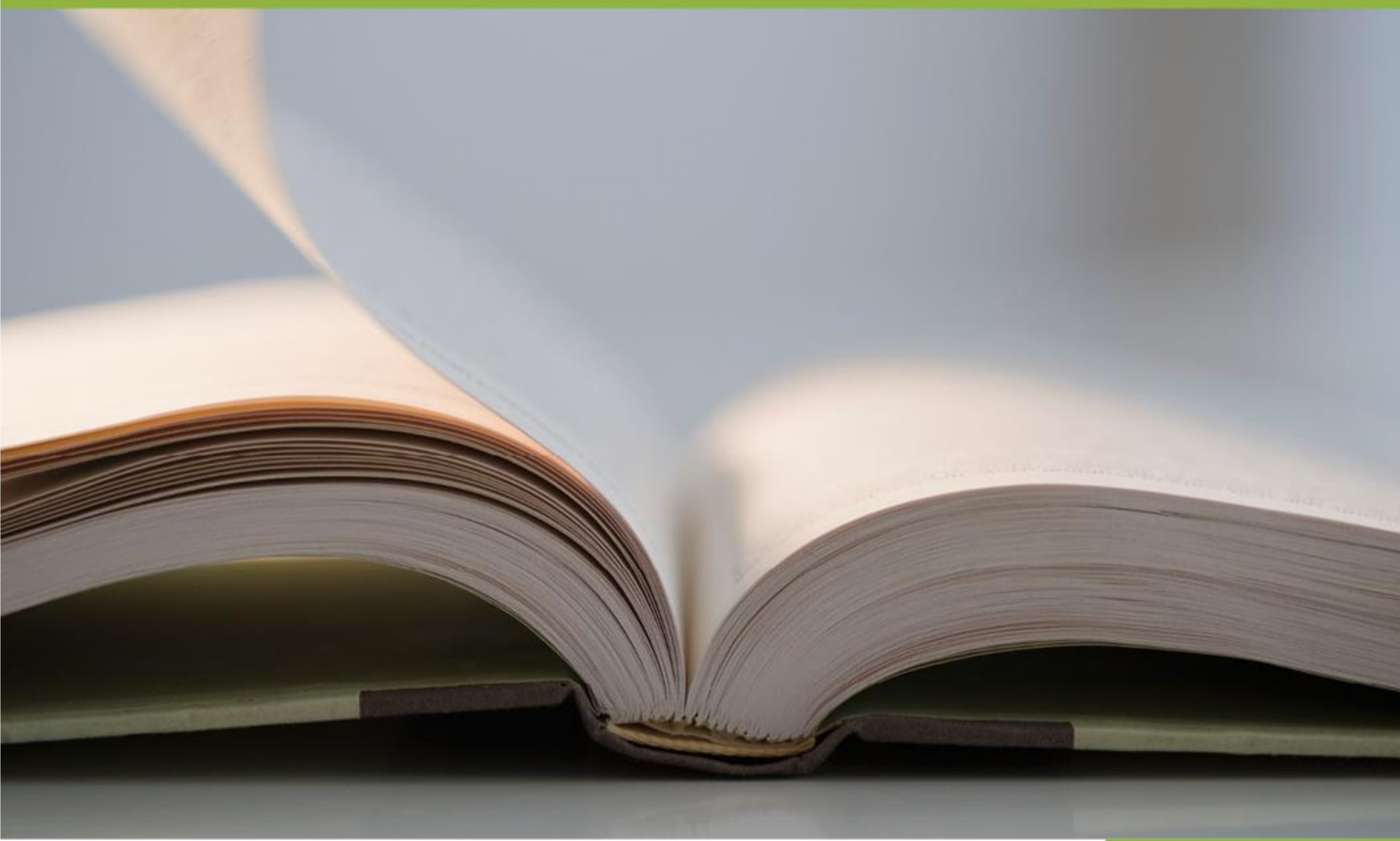
SAS® EVAAS

Statistical Models and Business Rules

Prepared for the Tennessee Department of Education



Contents

1 Introduction to Tennessee's Value-Added Reporting

Thirty years ago, the state of Tennessee led the nation in providing measures of student progress to individual districts, schools, and teachers. Known as the Tennessee Value-Added Assessment System (TVAAS), this reporting focused on the *progress* of students over time rather than their *proficiency level*. TVAAS represented a paradigm shift for educators and policymakers. In identifying the more effective practices and the less effective practices, educators receive personalized feedback, which they can then leverage to improve the academic experiences of their students.

TVAAS value-added reporting began with district reporting in 1993 and expanded to school reporting in 1994 and then teacher reporting in 1996.

The term "value-added" refers to a statistical analysis used to measure students' academic growth. Conceptually and as a simple explanation, value-added or growth measures are calculated by comparing the exiting achievement to the entering achievement for a group of students. Although the concept of growth is easy to understand, the implementation of a growth model is more complex.

First, there is not just one growth model; there are multiple growth models depending on the assessment, students included in the analysis, and level of reporting (district, school, or teacher). For each of these models, there are business rules to ensure the growth measures reflect the policies and practices selected by the State of Tennessee.

Second, in order to provide reliable growth measures, growth models must overcome non-trivial complexities of working with student assessment data. For example, students do not have the same entering achievement, students do not have the same set of prior test scores, and all assessments have measurement error because they are estimates of student knowledge.

Third, the growth measures are relative to students' expected growth, which is in turn determined by the growth that is observed within the actual population of Tennessee test-takers in a subject, grade, and year. Interpreting the growth measures in terms of their distance from expected growth provides a more nuanced, and statistically robust, interpretation.

With these complexities in mind, the purpose of this document is to guide you through Tennessee's value-added modeling based on the statistical models, business rules, policies, and practices selected by the State of Tennessee and currently implemented by EVAAS. This document includes details and decisions in the following areas:

- Conceptual and technical explanations of analytic models
- Definition of expected growth
- Classifying growth into categories for interpretation

- Explanation of district, school, and teacher composites
- Input data
- Business rules

These reports are delivered through the TVAAS web application available at http://tvaas.sas.com. Although the underlying statistical models and business rules supporting these reports are sophisticated and comprehensive, the web reports are designed to be user-friendly and visual so that educators and administrators can quickly identify strengths and opportunities for improvement and then use these insights to inform curricular, instructional, and planning supports.

2 Statistical Models

2.1 Overview of Statistical Models

The conceptual explanation of value-added reporting is simple: compare students' exiting achievement with their entering achievement over two points in time. In practice, however, measuring student growth is more complex. Students start the school year at different levels of achievement. Some students move around and have missing test scores. Students might have "good" test days or "bad" test days. Tests, standards, and scales change over time. A simple comparison of test scores from one year to the next does not incorporate these complexities. However, a more robust value-added model, such as the one used in Tennessee, can account for these complexities and scenarios.

Tennessee's value-added models offer the following advantages:

- **The models use multiple subjects and years of data.** This approach minimizes the influence of measurement error inherent in all academic assessments.
- **The models can accommodate students with missing test scores.** This approach means that more students are included in the model and represented in the growth measures. Furthermore, because certain students are more likely to have missing test scores than others, this approach provides less biased growth measures than growth models that cannot accommodate student with missing test scores.
- **The models can accommodate tests on different scales.** This approach gives flexibility to policymakers to change assessments as needed without a disruption in reporting. It permits more tests to receive growth measures, particularly those that are not tested every year.
- **The models can accommodate team teaching or other shared instructional practices.** This approach provides a more accurate and precise reflection of student learning among classrooms.

These advantages provide robust and reliable growth measures to districts, schools, and teachers. This means that the models provide valid estimates of growth given the common challenges of testing data. The models also provide measures of precision along with the individual growth estimates taking into account all of this information.

Furthermore, because this robust modeling approach uses multiple years of test scores for each student and includes students who are missing test scores, TVAAS value-added measures typically have very low correlations with student characteristics. It is not necessary to make direct adjustments for student socioeconomic status or demographic flags because each student serves as their own control. In other words, to the extent that background influences persist over time, these influences are already represented in the student's data. As a 2004 study by The Education Trust stated, specifically with regard to the EVAAS modeling:

> [I]f a student's family background, aptitude, motivation, or any other possible factor has resulted in low achievement and minimal learning growth in the past, all that is taken into account when the system calculates the teacher's contribution to student growth in the present.[1]

In other words, while technically feasible, adjusting for student characteristics in sophisticated modeling approaches is typically not necessary from a statistical perspective, and the value-added reporting in Tennessee does not make any direct adjustments for students' socioeconomic or demographic characteristics. Through this approach, the Tennessee Department of Education (TDOE) does not provide growth models to educators based on differential expectations for groups of students based on their backgrounds.

Based on Tennessee's state assessment program, there are two approaches to providing district, school, and teacher growth measures.

- **Gain model (also known as the multivariate response model or MRM)** is used for tests given in consecutive grades, such as Math and English Language Arts assessments in grades 4–8.
- **Predictive model (also known as univariate response model or URM)** is used when a test is given in non-consecutive grades or when performance from previous tests is used to predict performance on another test. This includes Math and English Language Arts in grade 3, Science in grades 5–8, Social Studies in grades 6–8, end-of-course (EOC) exams and ACT.

There is another model, which is similar to the predictive model except that it is intended as an instructional tool for educators serving students who have not yet taken an assessment.

- **Projection model** is used for all assessments and provides a probability of obtaining a particular score or higher on a given assessment for individual students.

The following sections provide technical explanations of the models. The online Help within the TVAAS web application is available at https://tvaas.sas.com, and it provides educator-focused descriptions of the models.

[1] Carey, Kevin. 2004. "The Real Value of Teachers: Using New Information about Teacher Effectiveness to Close the Achievement Gap." Thinking K-16 8(1):27.

2.2 Gain Model

2.2.1 Overview

The gain model measures growth between two points in time for a group of students; this is the case for tests given in consecutive grades such as Math and English Language Arts assessments in grades 4–8. **More specifically, the gain model measures the change in relative achievement for a group of students based on the statewide achievement from one point in time to the next.** For state summative assessments, growth is typically measured from one year to the next using the available consecutive grade assessments. Expected growth means that students maintained their relative achievement among the population of test-takers, and more details are available in Section 3.

There are three separate analyses for TVAAS reporting based on the gain model: one each for districts, schools, and teachers. The district and school models are essentially the same; they perform well with the large numbers of students in districts and most schools. The teacher model uses a version adapted to the smaller numbers of students typically found in teachers' classrooms.

In statistical terms, the gain model is known as a linear mixed model and can be further described as a multivariate repeated measures model. These models have been used for value-added analysis for almost three decades, but their use in other industries goes back much further. These models were developed to use in fields with very large longitudinal data sets that tend to have missing data.

Value-added experts consider the gain model to be among one of the most statistically robust and reliable models. The references below include foundational studies by experts from RAND Corporation, a non-profit research organization:

- On the **choice of a complex value-added model**: McCaffrey, Daniel F., and J.R. Lockwood. 2008. "Value-Added Models: Analytic Issues." Prepared for the National Research Council and the National Academy of Education, Board on Testing and Accountability Workshop on Value-Added Modeling, Nov. 13-14, 2008, Washington, DC.
- On the **advantages of the longitudinal, mixed model approach**: Lockwood, J.R. and Daniel McCaffrey. 2007. "Controlling for Individual Heterogeneity in Longitudinal Models, with Applications to Student Achievement." *Electronic Journal of Statistics* 1:223-252.
- On the **insufficiency of simple value-added models**: McCaffrey, Daniel F., B. Han, and J.R. Lockwood. 2008. "From Data to Bonuses: A Case Study of the Issues Related to Awarding Teachers Pay on the Basis of the Students' Progress." Presented at

Performance Incentives: Their Growing Impact on American K-12 Education, Feb. 28-29, 2008, National Center on Performance Incentives at Vanderbilt University.

2.2.2 Why the Gain Model is Needed

A common question is why growth cannot be measured with a simple gain model that measures the difference between the current year's scores and prior year's scores for a group of students. The example in Figure 1 illustrates why a simple approach is problematic.

Assume that 10 students are given a test in two different years with the results shown in Figure 1. The goal is to measure academic growth (gain) from one year to the next. Two simple approaches are to calculate the mean of the differences *or* to calculate the differences of the means. When there is no missing data, these two simple methods provide the same answer (5.8 on the left in Figure 1). When there is missing data, each method provides a different result (6.9 versus 4.6 on the right in Figure 1).

Figure 1: Scores without Missing Data, and Scores with Missing Data

Student	Previous Score	Current Score	Gain
1	51.9	74.8	22.9
2	37.9	46.5	8.6
3	55.9	61.3	5.4
4	52.7	47.0	-5.7
5	53.6	50.4	-3.2
6	23.0	35.9	12.9
7	78.6	77.8	-0.8
8	61.2	64.7	3.5
9	47.3	40.6	-6.7
10	37.8	58.9	21.1
Column Mean	**50.0**	**55.8**	**5.8**
Difference between Current and Previous Score Means			**5.8**

Student	Previous Score	Current Score	Gain
1	51.9	74.8	22.9
2		46.5	
3	55.9	61.3	5.4
4		47.0	
5	53.6	50.4	-3.2
6	23.0	35.9	12.9
7	78.6	77.8	-0.8
8	61.2	64.7	3.5
9	47.3	40.6	-6.7
10	37.8	58.9	21.1
Column Mean	**51.2**	**55.8**	**6.9**
Difference between Current and Previous Score Means			**4.6**

A more sophisticated model can account for the missing data and provide a more reliable estimate of the gain. As a brief overview, the gain model uses the correlation between current and previous scores in the nonmissing data to estimate means for all previous and current scores as if there were no missing data. It does this without explicitly imputing values for the missing scores. The difference between these two estimated means is an estimate of the average gain for this group of students. In this example, the gain model calculates the estimated difference to be 5.8. Even in a small example such as this, the estimated difference is much closer to the difference with no missing data than either measure obtained by the mean of the differences (6.9) or the difference of the means (4.6). This method of estimation has been shown, on average, to outperform both of the simple methods.[2] This small example only considered two grades and one subject for 10 students. Larger data sets, such as those used in the actual value-added analyses for the state, provide better correlation estimates by having more student data, subjects, and grades. In turn, these provide better estimates of means and gains.

[2] See, for example, S. Paul Wright, "Advantages of a Multivariate Longitudinal Approach to Educational Value-Added Assessment without Imputation," Paper presented at National Evaluation Institute, 2004. Available online at https://evaas.sas.com/support/EVAAS-AdvantagesOfAMultivariateLongitudinalApproach.pdf.

This simple example illustrates the need for a model that will accommodate incomplete data sets, which all student testing sets are. The next few sections provide more technical details about how the gain model calculates student growth.

2.2.3 Common Scale in the Gain Model

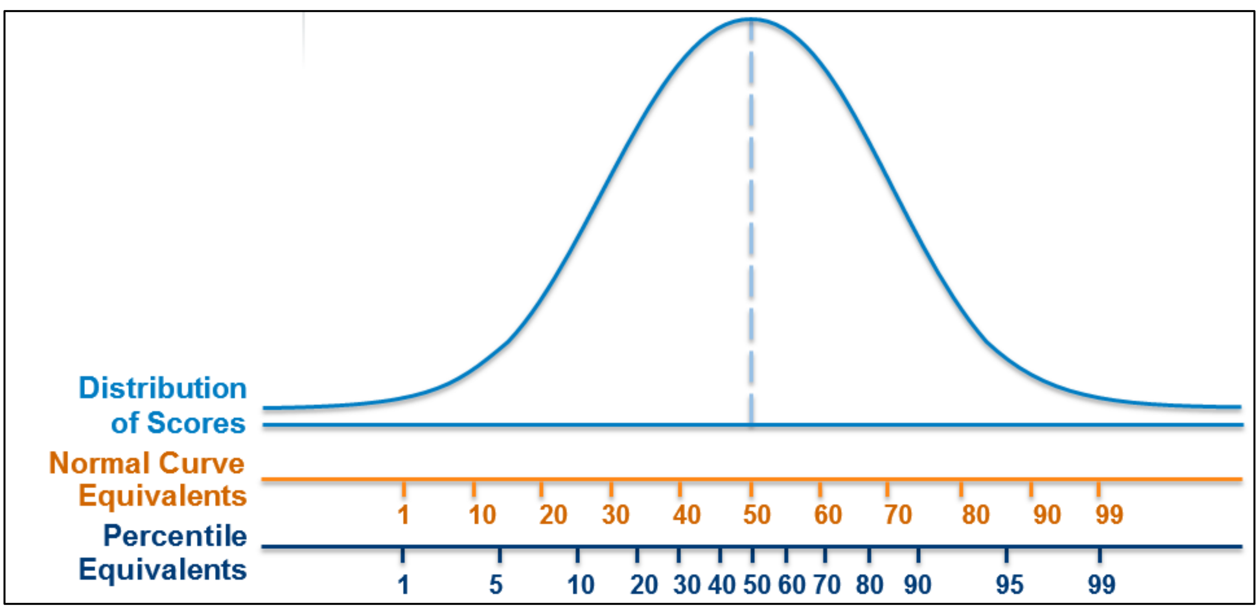
2.2.3.1 Why the Model Uses Normal Curve Equivalents

The gain model estimates academic growth as a "gain," or the difference between two measures of achievement from one point in time to the next. For such a difference to be meaningful, the two measures of achievement (that is, the two tests whose means are being estimated) must measure academic achievement on a common scale. Even for some vertically scaled tests, there can be different growth expectations for students based on their entering achievement. A reliable alternative whether tests are vertically scaled is to convert scale scores to normal curve equivalents (NCEs).

An NCE distribution is similar to a percentile one. Both distributions provide context as to whether a score is relatively high or low compared to the other scores in the distribution. In fact, NCEs are constructed to be equivalent to percentile ranks at 1, 50 and 99 and to have a mean of 50 and standard deviation of approximately 21.063.

However, NCEs have a critical advantage over percentiles for measuring growth: NCEs are on an equal-interval scale. This means that for NCEs, unlike percentile ranks, the distance between 50 and 60 is the same as the distance between 80 and 90. This difference between the distributions is evident below in Figure 2.

Figure 2: Distribution of Achievement: Scores, NCEs, and Percentile Rankings



Furthermore, percentile ranks are usually truncated below 1 and above 99, and NCEs can range below 0 and above 100 to preserve the equal-interval property of the distribution and to avoid truncating the test scale. In a typical year among Tennessee's state assessments, the average maximum NCE is approximately 120. The gain model does not use truncated values, which would create an artificial floor or ceiling in students' test scores.

Each NCE distribution is based on a specific assessment, test, subject, and time point. For example, the NCE distribution for 2024 TCAP Math in grade 5 is constructed separately from the NCE distribution for 2024 TCAP Math in grade 4.

2.2.3.2 Sample Scenario: How to Calculate NCEs in the Gain Model

The NCE distributions used in the gain model are based on a reference distribution of test scores in Tennessee. This reference distribution is the distribution of scores on a state-mandated test for all students in a given year. By definition, the mean (or average) NCE score for the reference distribution is 50 for each grade and subject. For identifying the other NCEs, the gain model uses a method that does not assume that the underlying scale is normal. This method ensures an equal-interval scale, even if the testing scales are not normally distributed.

Table 1 provides an example of how the gain model converts scale scores to NCEs. The first five columns of the table are based on a tabulated distribution of about 130,000 test scores from Tennessee data. In a given subject, grade, and year, the tabulation shows, for each given score, the number of students who scored that score ("Frequency") as well as the percentage ("Percent") that frequency represents out of the entire population of test-takers. The table also tabulates the "Cumulative Frequency" as the number of students who made that score or lower and its associated percentage ("Cumulative Percent").

The next column, "Percentile Rank," converts each score to a percentile rank. As a sample calculation using the data in Table 1 below, the score of 322 has a percentile rank of 45.2. The data show that 43.5% of students scored *below* 322 while 46.9% of students scored *at or below* 322. To calculate percentile ranks with discrete data, the usual convention is to consider half of the 3.4% reported in the "Percent" column to be "below" the cumulative percent and "half" above the cumulative percent. To calculate the percentile rank, half of 3.4% (1.7%) is added to 43.5% from "Cumulative Percent" to give you a percentile rank of 45.2, as shown in the table.

Table 1: Converting Tabulated Test Scores to NCE Values

Score	Frequency	Cumulative Frequency	Percent	Cumulative Percent	Percentile Rank	Z-Score	NCE
313	3,996	48,246	3.1	36.9	35.4	-0.375	42.10
315	4,265	52,511	3.3	40.2	38.5	-0.291	43.87
318	4,360	56,871	3.3	43.5	41.8	-0.206	45.66
322	4,404	61,275	3.4	46.9	45.2	-0.121	47.46
325	4,543	65,818	3.5	50.4	48.6	-0.035	49.27
328	4,619	70,437	3.5	53.9	52.1	0.053	51.12
330	4,645	75,082	3.6	57.4	55.7	0.143	53.00

NCEs are obtained from the percentile ranks using the normal distribution. The table of the standard normal distribution (found in many textbooks[3]) or computer software (for example, a spreadsheet) provides the associated Z-score from a standard normal distribution for any given percentile rank. NCEs are Z-scores that have been rescaled to have a "percentile-like" scale. As mentioned above, the NCE distribution is scaled so that NCEs exactly match the percentile ranks at 1, 50, and 99. To do this, each Z-score is multiplied by approximately 21.063 (the standard deviation on the NCE scale) and then 50 (the mean on the NCE scale) is added.

With the test scores converted to NCEs, growth is calculated as the difference from one year and grade to the next in the same subject for a group of students. This process is explained in more technical detail in the next section.

2.2.4 Technical Description of the Gain Model

2.2.4.1 Definition of the Linear Mixed Model

As a linear mixed model, the gain model for district, school, and teacher value-added reporting is represented by the following equation in matrix notation:

$$y = X\beta + Zv + \epsilon \tag{1}$$

y (in the growth context) is the $m \times 1$ observation vector containing test scores (usually NCEs) for all students in all academic subjects tested over all grades and years.

X is a known $m \times p$ matrix that allows the inclusion of any fixed effects.

[3] See, for example, the inside front cover of William Mendenhall, Richard L. Scheaffer, and Dennis D. Wackerly, *Mathematical Statistics with Applications* (Boston: Duxbury Press, 1986).

β is an unknown $p \times 1$ vector of fixed effects to be estimated from the data.

Z is a known $m \times q$ matrix that allows the inclusion of random effects.

v is a non-observable $q \times 1$ vector of random effects whose realized values are to be estimated from the data.

ϵ is a non-observable $m \times 1$ random vector variable representing unaccountable random variation.

Both v and ϵ have means of zero, that is, $E(v = 0)$ and $E(\epsilon = 0)$. Their joint variance is given by:

$$Var\begin{bmatrix} v \\ \epsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \tag{2}$$

where R is the $m \times m$ matrix that reflects the amount of variation in and the correlation among the student scores residual to the specific model being fitted to the data, and G is the $q \times q$ variance-covariance matrix that reflects the amount of variation in and the correlation among the random effects. If (v, ϵ) are normally distributed, the joint density of (y, v) is maximized when β has value b and v has value u given by the solution to the following equations, known as Henderson's mixed model equations:[4]

$$\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{bmatrix} \tag{3}$$

Let a generalized inverse of the above coefficient matrix be denoted by

$$\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{bmatrix}^{-} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = C \tag{4}$$

If G and R are known, then some of the properties of a solution for these equations are:

1. Equation (5) below provides the best linear unbiased estimator (BLUE) of the estimable linear function, $K^T \beta$, of the fixed effects. The second equation (6) below represents the variance of that linear function. The standard error of the estimable linear function can be found by taking the square root of this quantity.

$$E(K^T \beta) = K^T b \tag{5}$$

[4] McLean, Robert A., William L. Sanders, and Walter W. Stroup (1991). "A Unified Approach to Mixed Linear Models." The American Statistician, Vol. 45, No. 1, pp. 54-64.

$$Var(K^T b) = (K^T)C_{11}K \tag{6}$$

2. Equation (7) below provides the best linear unbiased predictor (BLUP) of v.

$$E(v|u) = u \tag{7}$$

$$Var(u - v) = C_{22} \tag{8}$$

where u is unique regardless of the rank of the coefficient matrix.

3. The BLUP of a linear combination of random and fixed effects can be given by equation (9) below provided that $K^T\beta$ is estimable. The variance of this linear combination is given by equation (10).

$$E(K^T\beta + M^T v\ |u) = K^T b + M^T u \tag{9}$$

$$Var\big(K^T(b - \ \beta) + M^T(u - v)\big) = (K^T M^T)C(K^T M^T)^T \tag{10}$$

4. With G and R known, the solution for the fixed effects is equivalent to generalized least squares, and if v and ϵ are multivariate normal, then the solutions for β and v are maximum likelihood.

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

6. If v and ϵ are not multivariate normal, then the solution to the mixed model equations still provides the maximum correlation between v and u.

2.2.4.2 District and School Models

The district and school gain models do not contain random effects; consequently, the Zv term drops out in the linear mixed model. The X matrix is an incidence matrix (a matrix containing only zeros and ones) with a column representing each interaction of school (in the school model), subject, grade, and year of data. The fixed-effects vector β contains the mean score for each school, subject, grade, and year with each element of β corresponding to a column of X. Since gain models are generally run with each school uniquely defined across districts, there is no need to include districts in the model.

Unlike the case of the usual linear model used for regression and analysis of variance, the elements of ϵ are not independent. Their interdependence is captured by the variance-covariance matrix, which is also known as the R matrix. Specifically, scores belonging to the same student are correlated. If the scores in y are ordered so that scores belonging to the same student are adjacent to one another, then the R matrix is block diagonal with a block, R_i, for each student. Each student's R_i is a subset of the "generic" covariance matrix R_0 that contains a row and column for each subject and grade. Covariances among subjects and grades are assumed to be the same for all years (technically, all cohorts), but otherwise the R_0 matrix is unstructured. Each student's R_i contains only those rows and columns from R_0 that match the subjects and grades for which the student has test scores. In this way, the gain model can use all available scores from each student.

Algebraically, the district gain model is represented as:

$$y_{ijkld} = \mu_{jkld} + \epsilon_{ijkld} \tag{11}$$

where y_{ijkld} represents the test score for the i^{th} student in the j^{th} subject in the k^{th} grade during the l^{th} year in the d^{th} district. μ_{jkld} is the estimated mean score for this particular district, subject, grade, and year. ϵ_{ijkld} is the random deviation of the i^{th} student's score from the district mean.

The school gain model is represented as:

$$y_{ijkls} = \mu_{jkls} + \epsilon_{ijkls} \tag{12}$$

This is the same as the district analysis with the addition of the subscript s representing s^{th} school.

The gain model uses multiple years of student testing data to estimate the covariances that can be found in the matrix R_0. This estimation of covariances is done within each level of analyses and can result in slightly different values within each analysis.

Solving the mixed model equations for the district or school gain model produces a vector b that contains the estimated mean score for each school (in the school model), subject, grade, and year. To obtain a value-added measure of average student growth, a series of computations can be done using the students from a school in a particular year and their prior and current testing data. The model produces means in each subject, grade, and year that can be used to calculate differences in order to obtain gains. Because students might change schools from one year to the next (in particular when transitioning from elementary to middle school, for example), the estimated mean score for the prior year/grade uses students who existed in the current year of that school. Therefore, mobility is taken into account within the model. Growth of students is computed using all students in each school including those that might have moved buildings from one year to the next.

The computation for obtaining a growth measure can be thought of as a linear combination of fixed effects from the model. The best linear unbiased estimate for this linear combination is given by equation (5). The growth measures are reported along with standard errors, and these can be obtained by taking the square root of equation (6) as described above.

2.2.4.3 Teacher Model

The teacher estimates use a more conservative statistical process to lessen the likelihood of misclassifying teachers. Each teacher's growth measure is assumed to be equal to the state average in a specific year, subject, and grade until the weight of evidence pulls them either above or below that state average. The model also accounts for the percentage of instructional responsibility the teacher has for each student during the school year. Furthermore, the teacher model is "layered," which means that:

- Students' performance with both their current and previous teacher effects are incorporated.
- For each school year, the teacher estimates are based students' testing data collected over multiple previous years.

Each element of the statistical model for teacher value-added modeling provides an additional level of protection against misclassifying each teacher estimate.

To allow for the possibility of many teachers with relatively few students per teacher, the gain model enters teachers as random effects via the Z matrix in the linear mixed model. The X matrix contains a column for each subject, grade, and year, and the b vector contains an estimated state mean score for each subject, grade, and year. The Z matrix contains a column for each subject, grade, year, and teacher, and the u vector contains an estimated teacher effect for each subject, grade, year, and teacher. The R matrix is as described above for the district or school model. The G matrix contains teacher variance components with a separate unique variance component for each subject, grade, and year. To allow for the possibility that a teacher might be very effective in one subject and very ineffective in another, the G matrix is constrained to be a diagonal matrix. Consequently, the G matrix is a block diagonal matrix with a block for each subject/grade/year. Each block has the form ${\sigma^2}_{jkl}I$ where ${\sigma^2}_{jkl}$ is the teacher variance component for the j^{th} subject in the k^{th} grade in the l^{th} year, and I is an identity matrix.

Algebraically, the teacher model is represented as:

$$y_{ijkl} = \mu_{jkl} + \left(\sum_{k^* \le k} \sum_{t=1}^{T_{ijk^*l^*}} w_{ijk^*l^*t} \times \tau_{jk^*l^*t} \right) + \epsilon_{ijkl} \tag{13}$$

y_{ijkl} is the test score for the i^{th} student in the j^{th} subject in the k^{th}grade in the l^{th} year. $\tau_{jk^*l^*t}$ is the teacher effect of the t^{th} teacher in the j^{th} subject in grade k^* in year l^*. The complexity of the parenthesized term containing the teacher effects is due to two factors. First, in any given subject, grade, and year, a student might have more than one teacher. The inner (rightmost) summation is over all the teachers of the i^{th} student in a particular subject, grade, and year, denoted by $T_{ijk^*l^*}$. $\tau_{jk^*l^*t}$ is the effect of the t^{th} teacher. $w_{ijk^*l^*t}$ is the fraction of the i^{th} student's instructional time claimed by the t^{th} teacher. Second, as mentioned above, this model allows teacher effects to accumulate over time. The outer (leftmost) summation accumulates teacher effects not only for the current (subscripts k and l) but also over previous grades and years (subscripts k^* and l^*) in the same subject. Because of this accumulation of teacher effects, this type of model is often called the "layered" model.

In contrast to the model for many district and school estimates, the value-added estimates for teachers are not calculated by taking differences between estimated mean scores to obtain mean gains. Rather, this teacher model produces teacher "effects" (in the u vector of the linear mixed model). It also produces state-level mean scores (for each year, subject, and grade) in the fixed-effects vector b. Because of the way the X and Z matrices are encoded, in particular because of the "layering" in Z, teacher gains can be estimated by adding the teacher effect to the state mean gain. That is, the interpretation of a teacher effect in this teacher model is as a gain expressed as a deviation from the average gain for the state in a given year, subject, and grade.

Table 2 illustrates how the Z matrix is encoded for three students (X, Y, and Z) who have three different scenarios of teachers during grades 3, 4, and 5 in two subjects, Math (M) and Reading (R). Teachers are identified by the letters A–F.

Student X's teachers represent the conventional scenario. Student X is taught by a single teacher in both subjects each year (teachers A, C, and E in grades 3, 4, and 5, respectively). Notice that in Student X's Z matrix rows for grade 4 there are ones (representing the presence of a teacher effect) not only for fourth-grade teacher C but also for third-grade teacher A. This is how the "layering" is encoded. Similarly, in the grade 5 rows, there are ones for grade 5 teacher E, grade 4 teacher C, and grade 3 teacher A.

Student Y is taught by two different teachers in grade 3: teacher A for Math and teacher B for Reading. In grade 4, Student Y had teacher C for Reading. For some reason, in grade 4 no teacher claimed Student Y for Math even though Student Y had a grade 4 Math test score. This score can still be included in the analysis by entering zeros into the Student Y's Z matrix rows for grade 4 Math. In grade 5, however, Student Y had no test score in Reading. This row is completely omitted from the Z matrix. There will always be a Z matrix row corresponding to each test score in the y vector. Since Student Y has no entry in y for grade 5 Reading, there can be no corresponding row in Z.

Student Z's scenario illustrates team teaching. In grade 3 Reading, Student Z received an equal amount of instruction from teachers A and B. The entries in the Z matrix indicate each teacher's contribution, 0.5 for each teacher. In grade 5 Math, however, Student Z was taught by both teachers E and F, but they did not make an equal contribution. Teacher E claimed 80% responsibility, and teacher F claimed 20%.

Because teacher effects are treated as random effects in this approach, their estimates are obtained by shrinkage estimation, which is technically known as best linear unbiased prediction or as empirical Bayesian estimation. This means that *a priori* a teacher is considered "average" (with a teacher effect of zero) until there is sufficient student data to indicate otherwise. This method of estimation protects against false positives (teachers incorrectly evaluated as most effective or least effective), particularly in the case of teachers with few students.

Table 2: Encoding the Z Matrix

			Teachers											
			Third Grade				Fourth Grade				Fifth Grade			
			A		B		C		D		E		F	
Student	**Grade**	**Subjects**	**M**	**R**	**M**	**R**	**M**	**R**	**M**	**R**	**M**	**R**	**M**	**R**
Student X	**3**	**M**	1	0	0	0	0	0	0	0	0	0	0	0
		R	0	1	0	0	0	0	0	0	0	0	0	0
	4	**M**	1	0	0	0	1	0	0	0	0	0	0	0
		R	0	1	0	0	0	1	0	0	0	0	0	0
	5	**M**	1	0	0	0	1	0	0	0	1	0	0	0
		R	0	1	0	0	0	1	0	0	0	1	0	0
Student Y	**3**	**M**	1	0	0	0	0	0	0	0	0	0	0	0
		R	0	0	0	1	0	0	0	0	0	0	0	0
	4	**M**	1	0	0	0	0	0	0	0	0	0	0	0
		R	0	0	0	1	0	1	0	0	0	0	0	0
	5	**M**	1	0	0	0	0	0	0	0	0	0	1	0
Student Z	**3**	**M**	1	0	0	0	0	0	0	0	0	0	0	0
		R	0	0.5	0	0.5	0	0	0	0	0	0	0	0
	4	**M**	1	0	0	0	0	0	1	0	0	0	0	0
		R	0	0.5	0	0.5	0	0	0	1	0	0	0	0
	5	**M**	1	0	0	0	0	0	1	0	0.8	0	0.2	0
		R	0	0.5	0	0.5	0	0	0	1	0	0	0	1

From the computational perspective, the teacher gain can be defined as a linear combination of both fixed effects and random effects and is estimated by the model using equation (9). The variance and standard error can be found using equation (10).

2.2.4.4 Student Groups Model

The gain model provides district and school growth measures for their students included in a specific student group. In this analysis, expected growth is the same as in the overall students' analysis. In other words, expected growth is based on all students since the NCE mapping is based on all students, not just those in a specific student group. Furthermore, the estimated covariance parameters are used from the overall students' analysis when calculating the value-added measures.

Students are identified as members of a student group based on a flag in the student record. Growth measures are calculated for each subset of students for each district and school that meet the minimum requirements of student data.

2.2.4.5 Lowest 25% Student Group

While most student groups are identified by information provided by TDOE, the lowest 25% designation is calculated by SAS for each grade and subject within each district and school. For grades and subjects using the gain model, students are identified based on the ranked average of their current and prior-year NCE values. The students with average scores in the lowest 25% of all scores at their school or district are identified as "Lowest 25%." This is conducted separately for each subject and grade within the school/district. Consequently, the identified students are not necessarily the same across subjects or between the district or school models. If the 25% threshold does not provide an integer, then the model will include the next lowest-score student so that the student group contains at least 25% of the school or district population.

Once the students are identified, there are additional steps to calculate growth for the lowest 25% student group for the school/district. When calculating the growth measures for the lowest 25% of students in a specific subject and grade, all the students' current and prior test scores across Math and Reading are included, just as they are in the overall school/district gain model. The same calculation is performed as in the overall school/district model to produce the growth measures for that group in the specific assessment. This is done separately for each subject, as each have a separate set of students. Finally, a combined Math and Reading model is run to create a composite standard error of any composite measures using components from this overall run and the separate Math and Reading models.

2.3 Predictive Model

2.3.1 Overview

Tests that are not given in consecutive grades require a different modeling approach from the gain model. The predictive model is used for such assessments in Tennessee. **The predictive model is a regression-based 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 with the average district, school, or teacher.

Like the gain model, there are three separate analyses for TVAAS reporting based on the predictive model: one each for districts, schools, and teachers. The district and school models are essentially the same, and the teacher model includes accommodations for team teaching and other shared instruction.

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

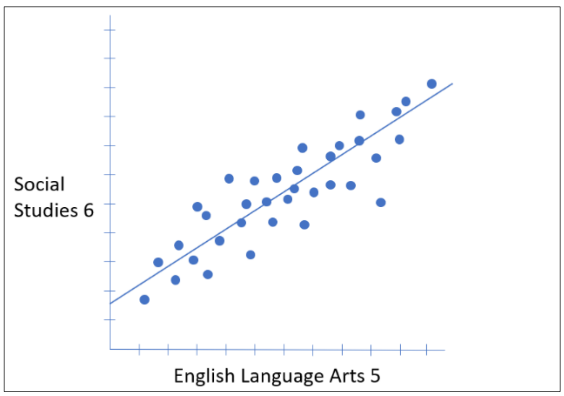
The key advantages of the predictive model can be summarized as follows:

- It minimizes the influence of measurement error and increases the precision of predictions by using multiple prior test scores as predictors for each student.
- It does not require students to have all predictors or the same set of predictors as long as a student has at least three prior test scores as predictors of the response variable in any subject and grade.
- It allows educators to benefit from all tests, even when tests are on different scales.
- It accommodates teaching scenarios where more than one teacher has responsibility for a student's learning in a specific subject, grade, and year.

2.3.2 Conceptual Explanation

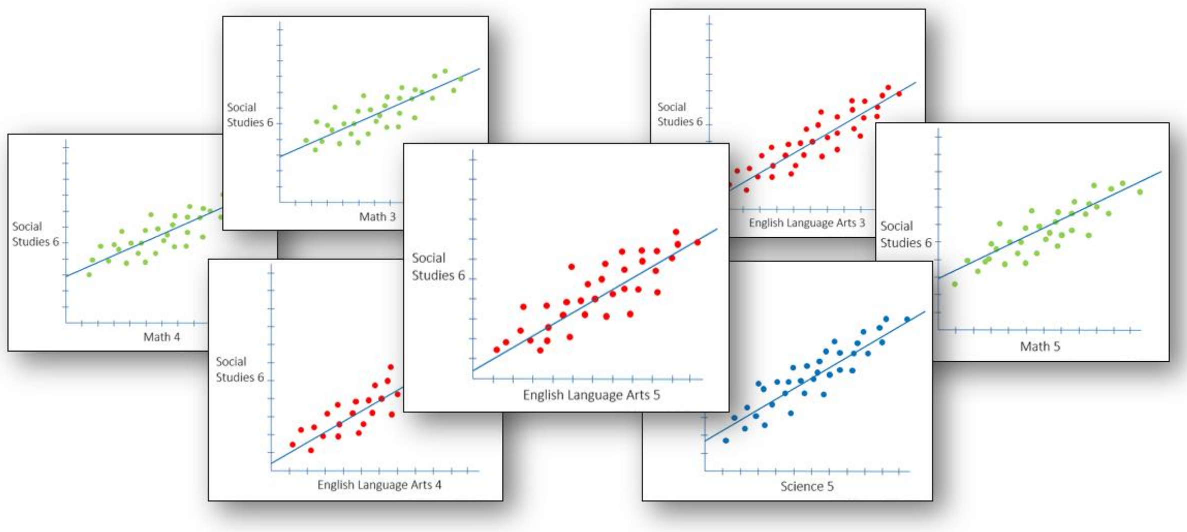
As mentioned above, the predictive model is ideal for assessments given in non-consecutive grades or when previous test performance is used to predict another test performance, such as TCAP Science in grades 5–8, Social Studies in grades 6–8, EOC, or ACT. Consider all students who tested in TCAP Social Studies in grade 6 in a given year. The gain model is not possible since there isn't a Social Studies test in the immediate prior grade. However, these students might have a number of prior test scores in TCAP Math and English Language Arts in grades 3–5 as well as TCAP Science in grade 5. These prior test scores have a relationship with TCAP Social Studies, meaning that how students performed on these tests can predict how the students perform on TCAP Social Studies in grade 6. The growth model does not assume what the predictive relationship will be; instead, the actual relationships observed in the data define the relationships. This is shown in Figure 3 below where each dot represents a student's prior score on TCAP English Language Arts 5 plotted with their score on TCAP Social Studies 6. The best-fit line indicates how students with a certain prior score on TCAP English Language Arts 5 tend to score, on average, on TCAP Social Studies 6. This illustration is based on one prior test; the predictive model uses many prior test scores from different subjects and grades.

Figure 3: Test Score from One Assessment Have a Predictive Relationship to Test Scores from Another Assessment



Some subjects and grades will have a greater relationship to TCAP Social Studies in grade 6 than others; however, the other subjects and grades still have a predictive relationship. For example, prior English Language Arts scores might have a stronger predictive relationship to TCAP Social Studies in grade 6 than prior Math scores, but how a student performs on the TCAP Math test typically provides an idea of how we might expect a student to perform on average on TCAP Social Studies. This is shown in Figure 4 below where different tests have a predictive relationship with TCAP Social Studies in grade 6. All of these relationships are considered together in the predictive model with some tests weighted more heavily than others based on the strength of their predictive relationship.

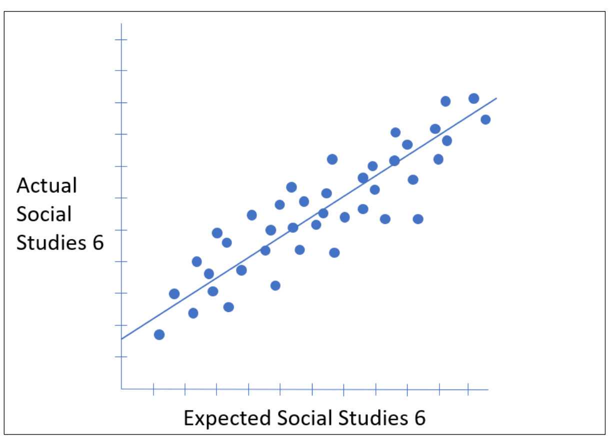
Figure 4: Relationships Observed in the Statewide Data Inform the Predictive Model



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

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. Figure 5 below shows the relationship between expected and actual scores for a group of students.

Figure 5: Relationship Expected Score and Actual Score for Selected Subject and Grade



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

2.3.3 Technical Description of the District, School, Teacher, and Student Groups Models

The predictive model has similar approaches for districts and schools and a slightly different approach for teachers that accounts for shared instructional responsibility. The approach is described briefly below, with more details following.

- The score to be predicted 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 taken in previous years from the response variable.
- There is a categorical variable (class variable, grouping variable) to identify the district, school, or teacher(s) 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, assuming in the teacher model that there is no team teaching.

$$y_i = \mu_y + \alpha_j + \beta_1(x_{i1} - \mu_1) + \beta_2(x_{i2} - \mu_2) + \cdots + \epsilon_i \quad (14)$$

In the case of team teaching, the single α_j is replaced by multiple α terms, each multiplied by an appropriate weight, similar to the way this is handled in the teacher gain model in equation (13). The μ terms are means for the response and the predictor variables. α_j is the teacher effect for the j^{th} district, school, or teacher—the one 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 and β terms). The parameter estimates (denoted with “hats,” e.g., $\hat{\mu}$, $\hat{\beta}$) are obtained using all students that have an observed value for the specific response and have three predictor scores. The resulting prediction equation for the i^{th} student is as follows:

$$\hat{y}_i = \hat{\mu}_y + \hat{\beta}_1(x_{i1} - \hat{\mu}_1) + \hat{\beta}_2(x_{i2} - \hat{\mu}_2) + \cdots \quad (15)$$

Two difficulties must be addressed in order to implement the prediction model. First, not all students will have the same set of predictor variables due to missing test scores. Second, because the predictive model is an ANCOVA model, the estimated parameters are pooled within group (district, school, or 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,

$$C = \begin{bmatrix} c_{yy} & c_{yx} \\ c_{xy} & C_{xx} \end{bmatrix} \quad (16)$$

Note that *C* in equation (16) is not the same as *C* in equation (4). This matrix is estimated using the expectation maximization (EM) algorithm for estimating covariance matrices in the presence of missing data available in SAS/STAT® (although no imputation is actually used). It should also be noted that, due to this being an ANCOVA model, *C* is a pooled-within group (district, school, or teacher) covariance matrix. This is accomplished by providing scores to the EM algorithm that are centered around group means (that is, the group means are subtracted from the scores) rather than around grand means. Obtaining *C* is an iterative process since group means are estimated within the EM algorithm to accommodate missing data. Once new group means are obtained, another set of scores is fed into the EM algorithm again until *C* converges. This overall iterative EM algorithm is what accommodates the two difficulties mentioned above. Only students who had a test score for the response variable in the most recent year and who had at least three predictor variables are included in the estimation. Given such a matrix, the vector of estimated regression coefficients for the projection equation (15) can be obtained as:

$$\hat{\beta} = C_{xx}^{-1} c_{xy} \quad (17)$$

This allows one to use 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 one imposes the restriction that the estimated "group" effects should sum to zero (that is, the effect for the "average" district, school or teacher is zero), then the appropriate means are the means of the group means. The group-level 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 group-level means.

Once the parameter estimates for the prediction equation have been obtained, predictions can be made using equation (15) for any student with any set of predictor values as long as that student has a minimum of three prior test scores. This is to avoid bias due to measurement error in the predictors.

The $\hat{y}_i$ term from equation (15) 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 achievement in the web reporting. The different prior test scores making up this composite are given different weights (by the regression coefficients, the $\hat{\beta}$ terms) 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 English Language Arts, 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 group effects (α_j) using the following ANCOVA model.

$$y_i = \gamma_0 + \gamma_1 \hat{y}_i + \alpha_j + \epsilon_i \tag{18}$$

In the predictive model, the effects (α_j) are considered random effects. Consequently, the $\hat{\alpha}_j$ terms are obtained by shrinkage estimation (empirical Bayes).[5] The regression coefficients for the ANCOVA model are given by the γ terms.

In the predictive model, there is an adjustment for Algebra I that considers the enrolled grade of the student, as there could be different enrolled grades for that assessment. This adjustment takes into account the relationship among student groups (those who take Algebra I in middle school versus those who take Algebra I in high school) so that there is neither an advantage nor a disadvantage to when students attributed to a district, school, or teacher take Algebra I.

2.3.3.1 Student Groups Model

In the analysis for specific student groups at a district or a school, expected growth is the same as in the overall students' analysis. In other words, expected scores ($\hat{y}$) are from the overall model and are the same as those used in the student group model. Furthermore, the estimated covariance parameters are used from the overall students' analysis when calculating the value-added measures in the context of shrinkage estimation.

2.3.3.2 Lowest 25% Student Group

While most student groups are identified by information provided by TDOE, the lowest 25% designation is calculated by SAS for each grade and subject within each district and school. For grades and subjects or courses using the predictive model, students are identified based their predicted score for a specific assessment as determined by the school/district-level predictive

---

[5] For more information about shrinkage estimation, see, for example, Ramon C. Littell, George A. Milliken, Walter W. Stroup, Russell D. Wolfinger, and Oliver Schabenberger, *SAS for Mixed Models, Second Edition* (Cary, NC: SAS Institute Inc., 2006). Another example is Charles E. McCulloch, Shayle R. Searle, and John M. Neuhaus, *Generalized, Linear, and Mixed Models, Second Edition* (Hoboken, NJ: John Wiley & Sons, 2008).

model. The students with predicted scores in the lowest 25% of all predicted scores at their school or district are identified as “Lowest 25%.” This is conducted separately for each subject and grade or course within the school/district. Consequently, the identified students are not necessarily the same across subjects/courses or between the district or school models. If the 25% threshold does not provide an integer, then the model will include the next lowest-score student so that the student group contains at least 25% of the school or district population.

Once the students are identified, there are additional steps to calculate growth for the Lowest 25% group for each school/district. First, for each subject and grade or course, residuals are created for individual students using the students’ predicted score as calculated from the overall (all students) school/district model. This residual is calculated as the difference between each student’s predicted and actual score, just as with the overall school/district model. This residual is calculated for each student with the lowest 25% designation in the subject and grade or course for which they have that designation. The average residual for the lowest 25% students within subject and grade or course is then shrunk back toward zero, similar to what was described for the overall school/district model but based on the number of students in the lowest 25% student group. Standard errors are also created as described for the overall school/district model using just this subset of students as well.

2.4 Projection Model

2.4.1 Overview

The longitudinal data sets used to calculate growth measures for groups of students can also provide individual student projections to future assessments. A projection is reported as a probability of obtaining a specific score or above on an assessment, such as a 70% probability of scoring Met Expectations or above on the next summative assessment. The probabilities are based on the students’ own prior testing history as well as how the cohort of students who just took the assessment performed. Projections are available for state assessments as well as to college readiness assessments.

Projections are useful as a planning resource for educators, and they can inform decisions around enrollment, enrichment, remediation, counseling, and intervention to increase students’ likelihood of future success.

2.4.2 Technical Description

The statistical model that is used as the basis for the projections is, in traditional terminology, an analysis of covariance (ANCOVA) model. This model is the same statistical model used in the predictive model applied at the school level described in Section 2.3.3. In the projection model, the score to be projected serves as the response variable (y), the covariates (x terms) are scores on tests the student has already taken, and the categorical variable is the school at which 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.

$$y_i = \mu_y + \alpha_j + \beta_1(x_{i1} - \mu_1) + \beta_2(x_{i2} - \mu_2) + \cdots + \epsilon_i \tag{19}$$

The μ terms are means for the response and the predictor variables. α_j is the school effect for the j^{th} school, the school attended by the i^{th} student. The β terms are regression coefficients. Projections to the future are made by using this equation with estimates for the unknown parameters (μ terms, β terms, sometimes α_j). The parameter estimates (denoted with "hats," e.g., $\hat{\mu}$, $\hat{\beta}$) are obtained using the most current data for which response values are available. The resulting projection equation for the i^{th} student is:

$$\hat{y}_i = \hat{\mu}_y \pm \hat{\alpha}_j + \hat{\beta}_1(x_{i1} - \hat{\mu}_1) + \hat{\beta}_2(x_{i2} - \hat{\mu}_2) + \cdots + \epsilon_i \tag{20}$$

The reason for the "±" before the $\hat{\alpha}_j$ term is that since the projection is to a future time, the school that the student will attend is unknown, so this term is usually omitted from the projections. This is equivalent to setting $\hat{\alpha}_j$ to zero, that is, to assuming that the student encounters the "average schooling experience" in the future.

Two difficulties must be addressed to implement the projections. First, not all students will have the same set of predictor variables due to missing test scores. Second, because this is an ANCOVA model with a school effect i, the regression coefficients must be "pooled-within-school" regression coefficients. The strategy for dealing with these difficulties is the same as described in Section 2.3.3 using equations (15), (16), and (17) and will not be repeated here.

Typically, the parameter estimates are based on the cohort of students who most recently took the assessment. Once the parameter estimates for the projection equation have been obtained, projections can be made for any student with any set of predictor values. To protect against bias due to measurement error in the predictors, projections are made only for students who have at least three available predictor scores or, for grades 3 and 4 only, students who have two predictors (Math and English Language Arts in grade 3). In addition to the projected score itself, the standard error of the projection is calculated ($SE(\hat{y}_i)$). Given a projected score and its standard error, it is possible to calculate the probability that a student will reach some specified benchmark of interest (b). Examples are the probability of scoring at least Met Expectations on a future grade-level test or the probability of scoring at least an established college readiness benchmark. The probability is calculated as the area above the benchmark cutoff score using a normal distribution with its mean equal to the projected score and its standard deviation equal to the standard error of the projected score as described below. Φ represents the standard normal cumulative distribution function.

$$Prob(\hat{y}_i \geq b) = \Phi\left(\frac{\hat{y}_i - b}{SE(\hat{y}_i)}\right) \tag{21}$$

2.5 Outputs from the Models

The outputs of the value-added model are available to Tennessee educators with user credentials in the TVAAS web application available at https://tvaas.sas.com/. Note that, for teachers working in multiple schools within the same district, the Teacher Value-Added Reports in the TVAAS web application are displayed in the school for which the teacher has the largest number of full-time effective (FTE) students. For teachers working in multiple districts, there is a Teacher Value-Added Report based on each individual district and displayed in that specific district's reporting in the TVAAS web application. In this instance, the teacher's evaluation composite would appear in the district for which the teacher has the largest number of FTE students.

2.5.1 Gain Model

The gain model is used for courses where students test in consecutive grade-given tests. As such, **the gain model uses TCAP English Language Arts and Math in grades 3–8 to provide district, school, and teacher growth measures in the following content areas:**

- TCAP English Language Arts in grades 4–8
- TCAP Math in grades 4–8

In addition to the mean scores and mean gain for an individual subject, grade and year, the gain model can also provide the following:

- Cumulative gains across grades (for each subject and year)
- Multi-year up to 3-average gains (for each subject and grade) (only available for teacher reporting)
- Composite gains across subjects

In general, these are all different forms of linear combinations of the fixed effects (and random effects for the teacher model), and their estimates and standard errors are computed in the same manner described above in equations (5) and (6) for district and school models and in equations (9) and (10) for the teacher model.

Collectively, the different models provide metrics for a variety of purposes within the State of Tennessee. They are summarized in the list below:

- District growth measures
  - Overall students
  - American Indian or Alaskan Native

  - Asian
  - Black or African American
  - Black/Hispanic/American Indian or Alaska Native Students
  - Economically Disadvantaged Students
  - English Learner (EL)
  - Hispanic
  - Lowest 25%
  - Native Hawaiian or Pacific Islander
  - Students with Disabilities (SWD)
  - Super Subgroup: Economically Disadvantaged, Students with Disabilities, EL Students, or Black/Hispanic/American Indian or Alaska Native Students
  - White
- School growth measures
  - Overall students
  - American Indian or Alaskan Native
  - Asian
  - Black or African American
  - Black/Hispanic/American Indian or Alaska Native Students
  - Economically Disadvantaged Students
  - English Learner (EL)
  - Hispanic
  - Lowest 25%
  - Native Hawaiian or Pacific Islander
  - Students with Disabilities (SWD)
  - Super Subgroup: Economically Disadvantaged, Students with Disabilities, EL Students, or Black/Hispanic/American Indian or Alaska Native Students
  - White
- Teacher growth measures based on linked students who meet inclusion criteria

Note that more details about district, school, and teacher composites across subjects, grades, and years are available in Section 5.

2.5.2 Predictive Model

The predictive model is used for courses where students test in non-consecutive grade-given tests. As such, **the predictive model provides growth measures for districts, schools, and teachers in the following content areas:**

- TCAP English Language Arts in grade 3 (Note: This is available only when there is sufficient grade 2 data in the current and prior year. Not all districts administer grade 2 assessments in the current and prior year.)
- TCAP Math in grade 3 (Note: This is available only when there is sufficient grade 2 data in the current and prior year. Not all districts administer grade 2 assessments in the current and prior year.)
- TCAP Science in grades 5–8
- TCAP Social Studies in grades 6–8
- EOC Algebra I
- EOC Algebra II
- EOC Biology I
- EOC English I
- EOC English II
- EOC Geometry
- EOC Integrated Math I
- EOC Integrated Math II
- EOC Integrated Math III
- EOC U.S. History

The predictive model also provides district- and school-level growth measures only in the following content areas, which is based on all students and does not include any student groups:

- ACT

In addition to the mean scores and growth measures for an individual subject, grade, and year, the predictive model can also provide multi-year average growth measures (up to three years) for each subject and grade or course. This multi-year average is only provided at the teacher level.

Collectively, the different models provide metrics for a variety of purposes within the State of Tennessee. They are summarized in the list below:

- District growth measures
  - Overall students
  - American Indian or Alaskan Native
  - Asian
  - Black or African American
  - Black/Hispanic/American Indian or Alaska Native Students
  - Economically Disadvantaged Students

- English Learner (EL)
- Hispanic
- Lowest 25%
- Native Hawaiian or Pacific Islander
- Students with Disabilities (SWD)
- Super Subgroup: Economically Disadvantaged, Students with Disabilities, EL Students, or Black/Hispanic/American Indian or Alaska Native Students
- White

- School growth measures
  - Overall students
  - American Indian or Alaskan Native
  - Asian
  - Black or African American
  - Black/Hispanic/American Indian or Alaska Native Students
  - Economically Disadvantaged Students
  - English Learner (EL)
  - Hispanic
  - Lowest 25%
  - Native Hawaiian or Pacific Islander
  - Students with Disabilities (SWD)
  - Super Subgroup: Economically Disadvantaged, Students with Disabilities, EL Students, or Black/Hispanic/American Indian or Alaska Native Students
  - White
- Teacher growth measures based on linked students who meet inclusion criteria

Note that more details about district, school and teacher composites across subjects, grades, and years are available in Section 5.

2.5.3 Projection Model

Projections are provided to future state assessments as well as college readiness assessments. More specifically, TCAP projections are provided only to a student's next tested grade-level TCAP assessments based on that student's most recent tested grade, such as projections to grade 5 for students who most recently tested in grade 4. EOC projections start with students who last tested in grade 5. Projections are made to the performance levels Approaching Expectations, Met Expectations, and Exceeded Expectations depending on the assessment, and the individual cut scores depend on each subject and grade. Grade 3 projections are available only to students with districts, schools or teachers that participated in the optional Grade 2 Assessment. Due to differences in business rules and methodology, the projections used to measure Adequate Growth in ELA might differ from the projections reported through the TVAAS web application. To summarize, the following TCAP and EOC projections are available for students who meet the reporting criteria:

- Math and English Language Arts in grade 3 (Note: This is available only for districts, schools or teachers that have Grade 2 assessments in the current year)
- Math and English Language Arts in grades 4–8
- Science in grades 5–8
- Social Studies in grades 6–8
- EOC Algebra I, Algebra II, Biology I, English I, English II, Geometry, Integrated Math I, Integrated Math II, Integrated Math III, and U.S. History

ACT projections start with students who last tested in grade 5, and they are made to various college benchmarks. These projections will be provided for the following subject areas:

- ACT Composite
- ACT English
- ACT Mathematics
- ACT Reading
- ACT Science/Reasoning

Advanced Placement (AP) projections start with students who last tested in grade 6, and they are made to levels 2, 3, and 4. These projections are available for the following subject areas:

- AP Biology
- AP English Language and Composition
- AP English Literature and Composition
- AP Human Geography
- AP Psychology
- AP Statistics
- AP United States Government
- AP United States History
- AP World History

3 Expected Growth

3.1 Overview

Conceptually, growth is simply the difference between students' entering and exiting achievement. As noted in Section 2, 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.

A more detailed explanation of expected growth and how it is calculated is useful for the interpretation and application of growth measures.

3.2 Technical Description

Both the gain and predictive models define 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. Both models define expected growth within a year. This means that expected growth is always relative to how students' achievement has changed in the most recent year of testing rather than a fixed year in the past.

More specifically, **in the gain model, expected growth means that students maintained the same relative position with respect to the statewide student achievement that year. In the predictive model, expected growth means that students with a district, school, or teacher made the same amount of growth as students with the average district, school, or teacher in the state for that same year, subject, and grade**.

For both models, the growth measures tend to be centered on expected growth every year with approximately half of the district/school/teacher estimates above zero and approximately half of the district/school/teacher estimates below zero.

A change in assessments or scales from one year to the next does not present challenges to calculating expected growth. Through the use of NCEs, the gain model converts any scale to a relative position, and the predictive model already uses prior test scores from different scales to calculate the expected score. When assessments change over time, expected growth is still based on the relative change in achievement from one point in time to another.

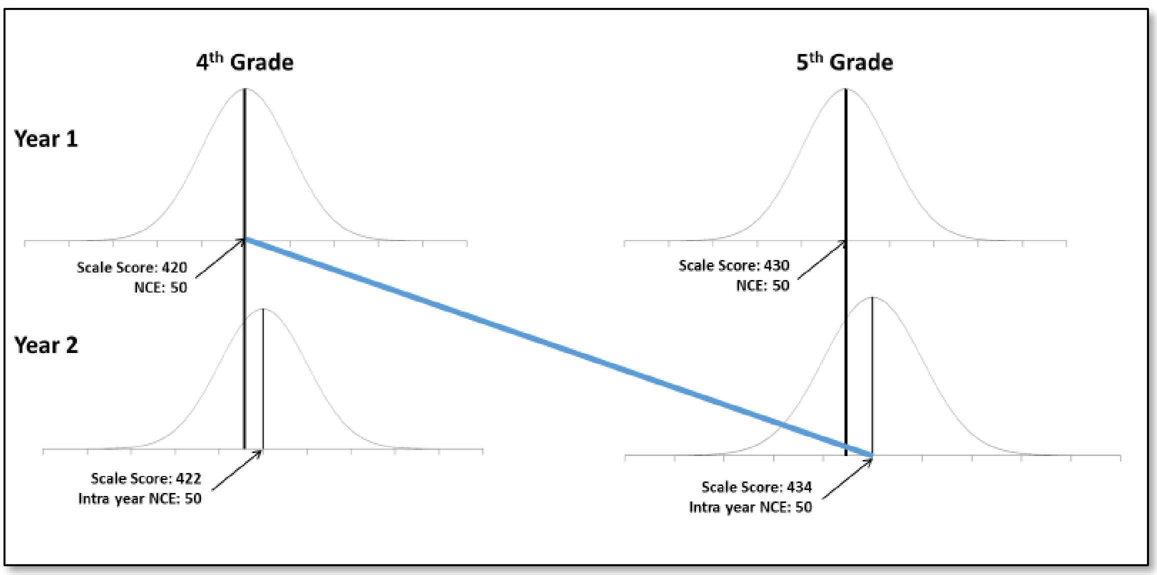
3.3 Illustrated Example

Figure 6 below provides a simplified example of how growth is calculated in the gain model when the state achievement increases. The figure has four graphs, each of which plot the NCE distribution of scale scores for a given year and grade. In this example, the figure shows how the gain is calculated for a group of grade 4 students in Year 1 as they become grade 5 students in

Year 2. In Year 1, our grade 4 students score, on average, 420 scale score points on the test, which corresponds to the 50th NCE (similar to the 50th percentile). In Year 2, the students score, on average, 434 scale score points on the test, which corresponds to a 50th NCE *based on the grade 5 distribution of scores in Year 2*. The grade 5 distribution of scale scores in Year 2 was higher than the grade 5 distribution of scale scores in Year 1, which is why the lower right graph is shifted slightly to the right. The blue line shows what is required for students to make expected growth, which would be to maintain their position at the 50th NCE for grade 4 in Year 1 as they become grade 5 students in Year 2. The growth measure for these students is Year 2 NCE – Year 1 NCE, which would be 50 – 50 = 0. Similarly, if a group of students started at the 35th NCE, the expectation is that they would maintain that 35th NCE.

Note that the actual gain calculations are much more robust than what is presented here; as described in the previous section, the models can address students with missing data, team teaching, and all available testing history.

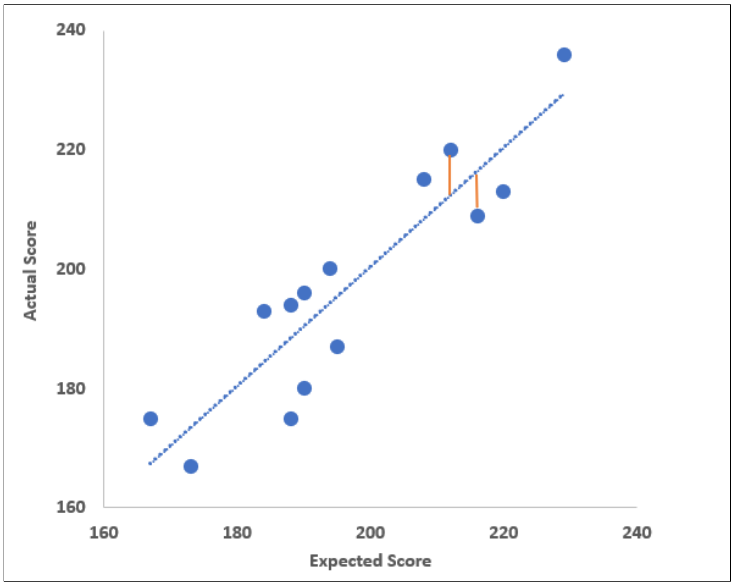
Figure 6: Intra-Year Approach Example for Gain Model



In contrast, in the predictive model, expected growth uses actual results from the most recent year of assessment data and considers the relationships from the most recent year with prior assessment results.

Figure 7 below provides a simplified example of how growth is calculated in the predictive model. The graph plots each student's actual score with their expected score. Each dot represents a student, and a best-fit line will minimize the difference between all students' actual and expected scores. Collectively, the best-fit line indicates what expected growth is for each student – given the student's expected score, expected growth is met if the student scores the corresponding point on the best-fit line. Conceptually, with the best-fit line minimizing the difference between all students' actual and expected scores, the growth expectation is defined by the average experience. Note that the actual calculations differ slightly since this is an ANCOVA model where the students are expected to see the average growth as seen by the experience with the average group (district, school, or teacher).

Figure 7: Intra-Year Approach Example for the Predictive Model



4 Classifying Growth into Categories

4.1 Overview

It can be helpful to classify growth into different levels for interpretation and context, particularly when the levels have statistical meaning. Tennessee's growth model has five categories for districts, schools, and teachers. These categories are defined by a range of values related to the growth measure and its standard error, and they are known as growth indicators in the web application.

4.2 Use Standard Errors Derived from the Models

As described in the modeling approaches section, the growth model provides an estimate of growth for a district, school, or teacher in a particular subject, grade, and year as well as that estimate's standard error. The standard error is a measure of the quantity and quality of student data included in the estimate, such as the number of students and the occurrence of missing data for those students. It also accounts for shared instruction and team teaching. Standard error is a common statistical metric reported in many analyses and research studies because it yields important information for interpreting an estimate, which is, in this case, the growth measure relative to expected growth. Because measurement error is inherent in any growth or value-added model, the standard error is a critical part of the reporting. **Taken together, the growth measure and standard error provide educators and policymakers with critical information about the certainty that students in a district, school, or classroom are making decidedly more or less than the expected growth.** Taking the standard error into account is particularly important for reducing the risk of misclassification (for example, identifying a teacher as ineffective when they are truly effective) for high-stakes usage of value-added reporting.

The standard error also takes into account that even among teachers with the same number of students, teachers might have students with very different amounts of prior testing history. Due to this variation, the standard errors in a given subject, grade, and year could vary significantly among teachers, depending on the available data that is associated with their students, and it is another important protection for districts, schools, and teachers to incorporate standard errors to the value-added reporting.

4.3 Define Growth Indicators in Terms of Standard Errors

Common statistical usage of standard errors indicates the precision of an estimate and whether that estimate is statistically significantly different from an expected value. The growth reports use the standard error of each growth measure to determine the statistical evidence that the growth measure is different from expected growth. For TVAAS growth reporting, this is essentially when the growth measure is more than or less than two standard errors above or below expected growth or, in other words, when the growth index is more than +2 or less than -2. These definitions then map to growth indicators in the reports themselves, such that there is statistical meaning in these categories. The categories and definitions are illustrated in the following section.

4.4 Illustrated Examples of Categories

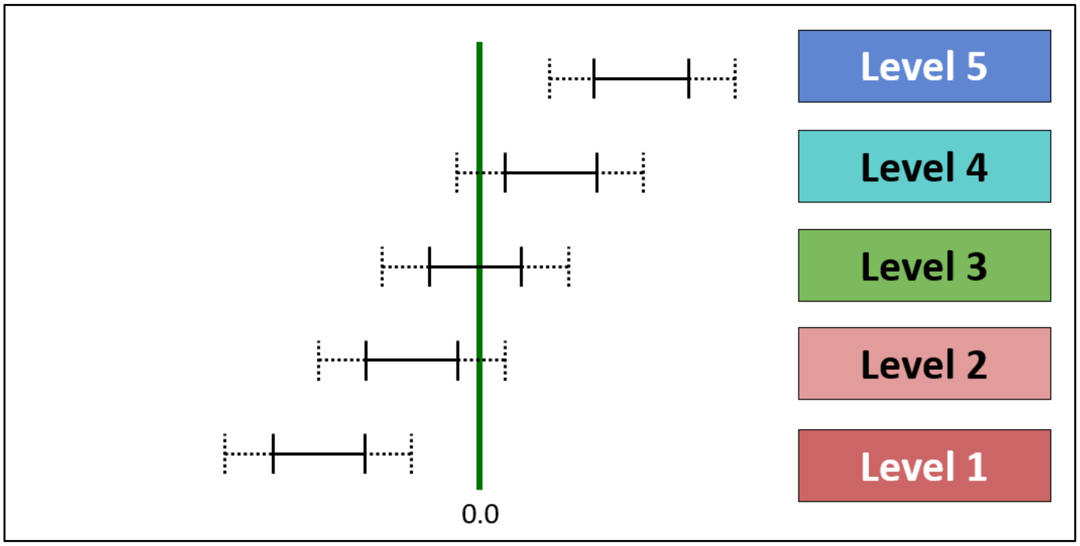
There are two ways to visualize how the growth measure and standard error relate to expected growth and how these can be used to create categories.

The first way is to frame the growth measure relative to its standard error and expected growth at the same time. For district and school reporting, the categories are defined as follows:

- **Level 5** indicates that the growth measure is two standard errors or more above expected growth (0). This level of certainty is significant evidence that students made more growth than expected.
- **Level 4** indicates that the growth measure is at least one but less than two standard errors above expected growth (0). This is moderate evidence that students made more growth than expected.
- **Level 3** indicates that the growth measure is less than one standard error above expected growth (0) but no more than one standard error below expected growth (0). This is evidence that students made growth as expected.
- **Level 2** indicates that the growth measure is more than one but no more than two standard errors below expected growth (0). This is moderate evidence that students made less growth than expected.
- **Level 1** is an indication that the growth measure is more than two standard errors below expected growth (0). This level of certainty is significant evidence that students made less growth than expected.

Figure 8 below shows visual examples of each category. The green line represents the expected growth. The solid black line represents the range of values included in the growth measure plus and minus *one* standard error. The dotted black line extends the range of values to the growth measure plus and minus *two* standard errors. If the dotted black line is completely above expected growth, then there is significant evidence that students made more than expected growth, which represents the Level 5 category. Conversely, if the dotted black line is completely below expected growth, then there is significant evidence that students made less than expected growth, which represents the Level 1 category. Levels 4 and 2 indicate, respectively, that there is moderate evidence that students made more than expected growth and less than expected growth. In these categories, the solid black line is completely above or below expected growth but not the dotted black line. Level 3 indicates that there is evidence that students made growth as expected as both the solid and dotted cross the line indicating expected growth.

Figure 8: Visualization of Growth Categories with Expected Growth, Growth Measures, and Standard Errors



This graphic is helpful in understanding how the growth measure relates to expected growth and whether the growth measure represents a statistically significant difference from expected growth.

The second way to illustrate the categories is to create a growth index, which is calculated as shown below:

$$Growth\ Index = \frac{Growth\ Measure - Expected\ Growth}{Standard\ Error\ of\ the\ Growth\ Measure} \quad (22)$$

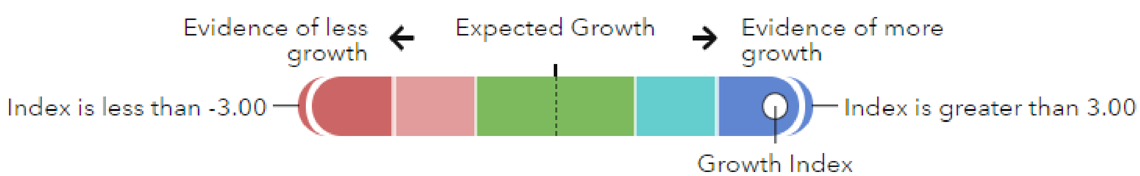
The growth index is similar in concept to a Z-score or t-value, and it communicates as a single metric the certainty or evidence that the growth measure is decidedly above or below expected growth. The growth index is useful when comparing value-added measures from different assessments or in different units, such as NCEs or scale scores. The categories can be established as ranges based on the growth index, such as the following:

- **Level 5** indicates significant evidence that students exceed the growth standard. The growth index is 2 or greater.
- **Level 4** indicates moderate evidence that students exceeded the growth standard. The growth index is between 1 and 2.
- **Level 3** indicates evidence that students met the growth standard. The growth index is between -1 and 1.

- **Level 2** indicates moderate evidence that students did not meet the growth standard. The growth index is between -2 and -1.
- **Level 1** indicates significant evidence that students did not meet the growth standard. The growth index is less than -2.

This is represented in the growth indicator bar in Figure 9, which is similar to what is provided in the District and School Value-Added Reports in the TVAAS web application. The black dotted line represents expected growth. The color-coding within the bar indicates the range of values for the growth index within each category.

Figure 9: Sample Growth Indicator Bar



It is important to note that these two illustrations provide users with the same information; they are simply presenting the growth measure, its standard error, and expected growth in different ways.

4.5 Effect Sizes

An additional measure that is helpful for interpreting growth is the effect size. While the growth index is the growth estimate divided by the standard error, the effect size is the growth measure divided by the student-level standard deviation of growth. The effect size provides an indicator of magnitude and practical significance that the group of students met, exceeded, or fell short of expected growth.

Effect sizes can be classified as small, medium, or large to assist with interpretation about the magnitude of measured growth compared to expectations. Various researchers have offered thoughts on what defines a small, medium, and large effect size. Based on research specific to education, Kraft suggested less than 0.05 as small, between 0.05 to 0.20 as medium, and greater than 0.20 as large based on the distributions of effect sizes and changes in achievement (Kraft MA. “Interpreting Effect Sizes of Education Interventions.” *Educational Researcher*. 2020; 49 (4):241-253).

Regardless of exact thresholds, researchers agree that it is important to interpret results within the distribution of actual results. In other words, what constitutes a small, medium, or large effect size is determined by what is observed in the actual results.

4.6 Rounding and Truncating Rules

As described in the previous section, the effectiveness level is based on the value of the growth index. As additional clarification, the calculation of the growth index uses unrounded values for the value-added measures and standard errors. After the growth index has been created but before the categories are determined, the index values are rounded or truncated by taking the maximum value of the rounded or truncated index value out to two decimal places. This provides the highest category given any type of rounding or truncating situation. For example, if the score was a 1.995, then rounding would provide a higher category. If the score was a -2.005, then truncating would provide a higher category. In practical terms, this impacts only a very small number of measures.

Also, when value-added measures are combined to form composites, as described in the next section, the rounding or truncating occurs after the final index is calculated for that combined measure.

5 Composite Growth Measures

A composite combines growth measures from different subjects, grades, and/or courses. The following sections provide information about Teacher and School Composites.

5.1 Teacher Composites

5.1.1 Overview

Teachers might receive evaluation composites based on their individual TVAAS value-added reporting, and teachers with a 2023-24 TVAAS teacher value-added measure are eligible to receive one or more of these composites. TDOE combines the TVAAS evaluation composites (growth measures) with qualitative measures and achievement measures to create a Level of Overall effectiveness (LOE) score for teachers.

For the 2023-24 reporting year, there are two evaluation composites possibly available for each teacher:

- **Single Year Composite,** which is comprised solely of value-added measures from 2023-24.
- **Multi-Year Composite (up to 3 years),** which is comprised of value-added measures from 2021-22, 2022-23 and 2023-24.

For each evaluation composite, the composite will include all available value-added measures within the years defined above for the teacher. This means the multi-year composite might not be comprised of measures from all three years if the teacher does not have any value-added measures for a particular year, so long as the teacher does have value-added measures for the 2023-24 school year. The value-added measures within the composite for a given year will be weighted according to the number of Full-Time Equivalent (FTE) students associated with each value-added measure. For multi-year composites, each year is then weighted equally.

5.1.2 Sample Calculation of Teacher Evaluation Composite

Table 3 below provides sample value-added measures for a teacher to illustrate how the evaluation composite is calculated.

Table 3: Sample Value-Added Measures for a Teacher

Year	Subject	Number of FTE Students	Value-Added Measure	Standard Error	Index
2024	Algebra I	25	15.50	5.50	2.82
2024	Algebra II	50	3.80	1.50	2.53
2024	Geometry	50	-0.30	1.20	-0.25
2023	Algebra I	25	3.47	1.60	2.17
2023	Algebra II	100	3.50	1.50	2.33
2022	Algebra II	50	2.80	1.30	2.15
2022	Geometry	25	0.40	1.10	0.36

Teacher evaluation composites could contain more than one scale since the various EOC assessments use different scales. Therefore, the value-added measures cannot simply be averaged across the seven different subject/grade/years for this sample teacher's evaluation composite. An index value can be used to combine them.

The index is standardized (unit-less) or in terms of the standard errors away from zero. This makes it possible to combine across subjects and grades. By definition and according to standard statistical theory, this standardized statistic has a standard error of 1.[6] The index is calculated for each teacher's value-added measure by dividing the value-added measure by its standard error. The index is reported in the final column of Table 3. As a reminder from earlier sections, the model produces a value-added measure and standard error for each year/subject/grade possible for a teacher. These two values are used to see whether there is statistical evidence that the value-added measure is different from the expectation of growth, which is zero.

5.1.3 Calculation of the Single-Year Evaluation Composites

To calculate the 2023-24 evaluation composite, the first step is to average the index values from the current year. In the above example, this would look like the following:

$$Unadjusted\ 2024\ Index = \left(\frac{25}{125} * 2.82 + \frac{50}{125} * 2.53 + \frac{50}{125} * (-0.25)\right) = 1.48 \quad (23)$$

Note that the index for each value-added measure is weighted according to the students associated with it. This teacher had 25 FTE students associated with the 2024 Algebra I value-

[6] See, for example, Dennis D. Wackerly, William Mendenhall III, and Richard L. Scheaffer, "Chapter 7" in *Mathematical Statistics with Applications, Sixth Edition* (Pacific Grove, CA: Duxbury Thomson Learning, Inc., 2002).

added measure, 50 FTE students associated with the 2024 Algebra II value-added measure, and 50 FTE students associated with the 2024 Geometry value-added measure. The total number of FTE students totals 25 + 50 + 50, or 125. The index for 2024 Algebra I (2.82) is thus weighted proportionately at 25/125, the index for 2024 Algebra II (2.53) is also weighted at 50/125, and the index for 2024 Geometry (-0.25) is weighted at 50/125. In equation (23) above and all other evaluation composite calculations, the unrounded index values are used (meaning, the value-added measure divided by its standard error rather than the rounded value reported in Table 3).

Since each of the individual index values have a standard error of 1, there needs to be an additional correction to recalculate the overall average index to make it have a standard error of 1 or so that it is standardized like the original index values. This standard error of an average index can be found using the following formula:

$$SE\ for\ 2024\ Index = \sqrt{\left(\frac{25}{125}\right)^2 + \left(\frac{50}{125}\right)^2 + \left(\frac{50}{125}\right)^2} = 0.60 \tag{24}$$

To calculate the new index, the average of the index values would be divided by the new standard error of the average index.

$$Final\ 2024\ Index = \frac{1.48}{0.60} = 2.46 \tag{25}$$

Notice how the index value of the composite is larger than the average index. This is because there is more information and evidence about students' growth when all the individual measures are combined. The additional evidence provides a greater level of certainty that this teacher's students are demonstrating above average growth across the subjects and grades in the current year.

5.1.4 Calculation of the Multi-Year Composites (up to 3 years)

The Multi-Year Composite includes up to three years of value-added measures based on the 2021-22, 2022-23 and 2023-24 reporting. To calculate this composite, the single-year composites for 2024 calculated in the previous sections would be combined with composites calculated similarly for 2022 and 2023, and each available yearly composite would be weighted equally.

To create the Multi-Year Composite, the 2021-22 and 2022-23 composites would first be calculated with the same steps from 5.1.3 using the 2022 and 2023 data listed in Table 3. Based on this information, the index for 2022 Algebra II (2.15) is weighted proportionately at 50/75, and the index for 2022 Geometry (0.36) is weighted at 25/75. The calculation of the 2022 index is as follows:

$$Unadjusted\ 2022\ Index = \left(\frac{50}{75} * 2.15 + \frac{25}{75} * 0.36\right) = 1.55 \tag{32}$$

Before combining the individual years into a multi-year index, each year's index is adjusted by dividing by the standard error as in the single year composite. The standard error for the 2022 unadjusted index value is 0.75. This is calculated in the same way as was done for the 2024 single-year composite:

$$SE\ for\ 2022\ Index = \sqrt{\left(\frac{50}{75}\right)^2 + \left(\frac{25}{75}\right)^2} = 0.75 \tag{33}$$

$$Final\ 2022\ Index = \frac{1.55}{0.75} = 2.08 \tag{34}$$

Similarly, for calculating the 2023 index, the index for 2023 Algebra I (2.17) is weighted proportionately at 25/125, and the index for 2023 Algebra II (2.33) is weighted at 100/125. The index is then adjusted by the standard error (0.82). The calculations for each step of creating the 2023 index are as follows:

$$Unadjusted\ 2023\ Index = \left(\frac{25}{125} * 2.17 + \frac{100}{125} * 2.33\right) = 2.30 \tag{26}$$

$$SE\ for\ 2023\ Index = \sqrt{\left(\frac{25}{125}\right)^2 + \left(\frac{100}{125}\right)^2} = 0.82$$

$$Final\ 2023\ Index = \frac{2.30}{0.82} = 2.79 \tag{28}$$

The next step is to calculate a multi-year index that combines the 2022, 2023, and 2024 indices according to their specified weights. Similar to the single year composites, this index is "unadjusted" and is not considered final until it is divided by its standard error.

$$Unadjusted\ 2022, 2023\ and\ 2024\ Index = \left(\frac{1}{3} * 2.08 + \frac{1}{3} * 2.79 + \frac{1}{3} * 2.46\right) = 2.44 \tag{35}$$

The standard error can again be calculated using the following formula, which accounts for the different weights of each year's index value in the overall multi-year index.

$$SE\ for\ 2022, 2023\ and\ 2024\ \ Index = \sqrt{\left(\frac{1}{3}\right)^2 + \left(\frac{1}{3}\right)^2 + \left(\frac{1}{3}\right)^2}\ =\ 0.58 \tag{36}$$

The new index value for the 2022, 2023 and 2023 reporting would be as follows (using non-rounded numbers):

$$Final\ 2022, 2023\ and\ 2024\ Index = \frac{2.44}{0.58} = 4.23 \tag{37}$$

5.2 District and School Evaluation Composites

Districts and schools also receive evaluation composites. The TDOE policies for these composites are outlined below:

- District and school evaluation composites are single-year measures based entirely on the current year's reporting.
- District and school evaluation composites weigh the value-added measures that are included in the composite according to the number of students associated with each value-added measure.
- There are six types of evaluation composites: Overall, Literacy, Numeracy, a combined Literacy and Numeracy, Science, and Social Studies. These six types can be created using different combinations of test data, and all options are listed in Section 5.2.7. Where applicable, the grades associated with each subject are included in parentheses.

5.2.1 Sample Calculation of District and School Evaluation Composite

Like Section 5.1, this section presents how school composites are calculated and how the decisions for schools share the same statistical approaches and policy decisions as those for teachers.

The key steps for determining a school's composite index are as follows:

1. For measures based on the gain model, calculate composite gain, standard error, and index across subjects and grades.
2. For measures based on the predictive model, calculate composite index across subjects.
3. Calculate composite index using both the gain and predictive model composite indices.

The following sections illustrate this process using value-added measures from a sample middle school, which are provided in Table 4 below.

Table 4: Sample School Value-Added Information

Year	Subject	Grade	Value-Added Gain	Standard Error	Number of Students
2024	Math	6	3.30	0.70	44
2024	ELA	6	-1.10	1.00	46
2024	Math	7	2.00	0.50	50
2024	ELA	7	2.40	1.10	50
2024	Math	8	-0.30	0.60	40
2024	ELA	8	3.80	0.70	50
2024	Algebra I	N/A	-11.50	6.20	35

5.2.2 For Gain Model Measures, Calculate Composite Gain Across Subjects

When the value-added estimates are in the same scale (Normal Curve Equivalents), the school composite gain across the six subject/grades is a weighted average based on the number of students in each subject and grade. For the school, the total number of students affiliated with gain value-added measures is 44 + 46 + 50 + 50 + 40 + 50, or 280. The Math grade 6 value-added measure would be weighted at 44/280, the ELA grade 6 value-added measure would be weighted at 46/280, and so on. More specifically, the composite gain is calculated using the following formula:

$$\begin{aligned} Comp\ Gain = {} & \tfrac{44}{280} Math_6 + \tfrac{46}{280} ELA_6 + \tfrac{50}{280} Math_7 + \tfrac{50}{280} ELA_7 + \tfrac{40}{280} Math_8 + \tfrac{50}{280} ELA_8 \\ = {} & \left(\tfrac{44}{280}\right)(3.30) + \left(\tfrac{46}{280}\right)(-1.10) + \left(\tfrac{50}{280}\right)(2.00) + \left(\tfrac{50}{280}\right)(2.40) + \left(\tfrac{40}{280}\right)(-0.30) + \left(\tfrac{50}{280}\right)(3.80) = 1.76 \end{aligned} \tag{38}$$

5.2.3 For Gain Model Measures, Calculate Standard Error Across Subjects

5.2.3.1 Technical Background on Standard Errors

The standard error of the school composite value-added gain cannot be calculated using the assumption that the gains making up the composite are independent when these individual measures are created with the gain model. This is because many of the same students are likely represented in different value-added gains, such as grade 8 Math in 2024 and grade 8 ELA in 2024. The statistical approach, outlined in Section 2.2.4 (with references), is quite sophisticated and will consider the correlations between pairs of value-added gains as shown in equation (39) below and using equation (12) for schools and equation (13) for teachers.[7] The composites are indeed linear combinations of the fixed effects of the models and can be estimated as described in Section 2.2.4. The magnitude of each correlation depends on the extent to which the same students are in both estimates for any two subject/grade/year estimates.

5.2.3.2 Illustration of Gain Model-Based Standard Error for Sample School

As a reminder, the use of the word "error" does not indicate a mistake. Rather, value-added models produce *estimates*. The value-added gains in the above tables are estimates, based on student test score data, of the school's true value-added effectiveness. In statistical terminology a "standard error" is a measure of the uncertainty in the estimate, providing a means to determine whether an estimate is decidedly above or below the growth expectation. Standard errors can and should also be provided for the composite gains that have been calculated, as shown above, from a teacher's value-added gain estimate.

Statistical formulas are often more conveniently expressed as variances, and this is the square of the standard error. Standard errors of composites can be calculated using variations of the general formula shown below. To maintain the generality of the formula, the individual estimates in the formula (think of them as value-added-gains) are simply called X, Y, and Z. If there were more than or fewer than three estimates, the formula would change accordingly. As gain model composites use proportional weighting according to the number of students linked to each value-added gain, each estimate is multiplied by a different weight: a, b, or c.

$$\begin{gathered} Var(aX + bY + cZ) = a^2Var(X) + b^2Var(Y) + c^2Var(Z) \\ +2ab\ Cov(X,Y) + 2ac\ Cov(X,Z) + 2bc\ Cov(Y,Z) \end{gathered} \tag{39}$$

Covariance, denoted by $Cov,$ is a measure of the relationship between two variables. It is a function of a more familiar measure of relationship, the correlation coefficient. Specifically, the term $Cov(X,Y)$ is calculated as follows:

$$Cov(X,Y) = Correlation(X,Y)\sqrt{Var(X)}\sqrt{Var(Y)} \tag{40}$$

[7] For more details about the statistical approach to derive the standard errors, see, for example, Ramon C. Littell, George A. Milliken, Walter W. Stroup, Russell D. Wolfinger, and Oliver Schabenberger, *SAS for Mixed Models, Second Edition* (Cary, NC: SAS Institute Inc., 2006). Another example: Charles E. McCulloch, Shayle R. Searle, and John M. Neuhaus, *Generalized, Linear, and Mixed Models* (Hoboken, NJ: Wiley, 2008).

The value of the correlation ranges from -1 to +1, and these values have the following meanings:

- A value of zero indicates no relationship.
- A positive value indicates a positive relationship, or Y tends to be larger when X is larger.
- A negative value indicates a negative relationship, or Y tends to be smaller when X is larger.

Two variables that are unrelated have a correlation and covariance of zero. Such variables are said to be statistically independent. If the X and Y values have a positive relationship, then the covariance will also be positive. As a general rule, two value-added gain estimates are statistically independent if they are based on completely different sets of students.

For our sample school's composite gain, the relationship will generally be positive. This means that the gain model-based composite standard error is larger than it would be assuming independence. Using the student weightings and standard errors reported in Table 4**Error! Reference source not found.** and assuming total independence, the standard error would then be as follows:

$$\begin{aligned} Comp\ Standard\ Error \ &= \sqrt{\begin{array}{l} \left(\frac{44}{280}\right)^2 (SE\ Math_6)^2 + \left(\frac{46}{280}\right)^2 (SE\ ELA_6)^2 + \left(\frac{50}{280}\right)^2 (SE\ Math_7)^2 \\ + \left(\frac{50}{280}\right)^2 (SE\ ELA_7)^2 + \left(\frac{40}{280}\right)^2 (SE\ Math_8)^2 + \left(\frac{50}{280}\right)^2 (SE\ ELA_8)^2 \end{array}} \\ &= \sqrt{\begin{array}{l} \left(\frac{44}{280}\right)^2 (0.70)^2 + \left(\frac{46}{280}\right)^2 (1.00)^2 + \left(\frac{50}{280}\right)^2 (0.50)^2 \\ + \left(\frac{50}{280}\right)^2 (1.10)^2 + \left(\frac{40}{280}\right)^2 (0.60)^2 + \left(\frac{50}{280}\right)^2 (0.70)^2 \end{array}} = 0.33 \end{aligned} \tag{41}$$

At the other extreme, if the correlation between each pair of value-added gains had its maximum value of +1, the standard error would be larger.

The actual standard error will likely be above the value of 0.33 due to students being in both Math and ELA in the school with the specific value depending on the values of the correlations between pairs of value-added gains. The magnitude of each correlation depends on the relationship between the subject and grades and the extent to which the same students are in both estimates for any two subject/grade estimates within a year.

For the sake of simplicity, let us assume the actual standard error was 0.40 for the school composite in this example.

5.2.4 For Gain Model Measures, Calculate Composite Index Across Subjects

The next step is to calculate the gain model-based school composite index, which is the school composite value-added gain divided by its standard error. The composite index for this school would be calculated as follows:

$$Composite\ Index = \frac{Composite\ Gain}{Composite\ Standard\ Error} = \frac{1.76}{0.40} = 4.40 \qquad (42)$$

Although some of the values in the example were rounded for display purposes, the actual rounding or truncating occurs only after all of measures have been combined as described in Section 4.6.

5.2.5 For Predictive Model Measures, Calculate Index Across Subjects

For our sample school, there is only one available value-added measure from the predictive model. This means that the reported value-added index for that subject will be the same that is calculated for the predictive model-based composite index.

$$Composite\ Index = \frac{Alg\ I\ Growth\ Measure}{Alg\ I\ Standard\ Error} = \frac{-11.50}{6.20} = -1.85 \qquad (43)$$

However, should a school or district have more than one value-added measure based on the predictive model, then the composite index would be calculated by first calculating index values for each subject and then combining those weighting by the number of students. The standard error of this combined index must assume independence since the predictive model measures are done in separate models for each year and subject.

5.2.6 Calculate the Combined Gain and Predictive Model Composite Index Across Subjects

The two composite indices from the gain and predictive models are then weighted according to the number of students within each model to determine the combined composite index. Our sample school has 315 students, of which 280 are in the gain model and 35 in the predictive model. The combined composite index would be calculated as follows using these weightings, the gain model-based composite index across subjects, and the predictive model-based index across subjects:

$$Unadjusted\ Combined\ Composite\ Index = \left(\frac{280}{315}\right)4.40 + \left(\frac{35}{315}\right)(-1.85) = 3.71 \qquad (44)$$

This combined index is not an actual index itself until it is adjusted to accommodate for the fact that it is based on multiple pieces of evidence together. An index, by definition, has a standard error of 1, but this unadjusted value (3.71) does not have a standard error of 1. The next step is to calculate the new standard error and divide the combined composite index found above by it. This new, adjusted composite index will be the final index with a standard error of 1. The standard error can be found given the standard formula above and the fact that each index has a standard error of 1. Independence is assumed since these are done outside of the models. In this example, the standard error would be as follows:

$$Final\ Combined\ Comp\ Standard\ Error = \sqrt{\left(\frac{280}{315}\right)^2 (1)^2 + \left(\frac{35}{315}\right)^2 (1)^2} = 0.90 \tag{45}$$

Therefore, the final combined composite index value is 3.71 divided by 0.90, or 4.14. This is the value that determines the school evaluation composite. Different types of evaluation composites use the value-added measures from different tests, but the overall process is the same.

5.2.7 Types of Evaluation Composites

5.2.7.1 Early Grades (Grade 3)

Composite Type	**Subjects**
Overall	Math (3), ELA (3)
Literacy	ELA (3)
Numeracy	Math (3)
Literacy and Numeracy	Math (3), ELA (3)
Science	N/A
Social Studies	N/A

5.2.7.2 TCAP (Grades 4-8)

Composite Type	**Subjects**
Overall	TCAP Math and English Language Arts (4–8), Science (5–8), and Social Studies (6–8)
Literacy	TCAP English Language Arts (4–8)
Numeracy	TCAP Math (4–8)
Literacy and Numeracy	TCAP Math and English Language Arts (4–8)
Science	Science (5-8)
Social Studies	Social Studies (6-8)

5.2.7.3 TCAP (Grades 4-8)/EOC

Composite Type	Subjects
Overall	Algebra I, Algebra II, Biology, English I, English II, Geometry, Integrated Math I, Integrated Math II, Integrated Math III, U.S. History, TCAP Math and English Language Arts (4–8), Science (5–8), and Social Studies (6–8)
Literacy	English I, English II, and TCAP English Language Arts (4–8)
Numeracy	Algebra I, Algebra II, Geometry, Integrated Math I, Integrated Math II, Integrated Math III, and TCAP Math (4–8)
Literacy and Numeracy	Algebra I, Algebra II, English I, English II, Geometry, Integrated Math I, Integrated Math II, Integrated Math III, and TCAP Math and English Language Arts (4–8)
Science	Biology and Science (5–8)
Social Studies	U.S. History and Social Studies (6-8)

5.2.7.4 EOC

Composite Type	Subjects
Overall	Algebra I, Algebra II, Biology, English I, English II, Geometry, Integrated Math I, Integrated Math II, Integrated Math III, and U.S. History
Literacy	English I, English II
Numeracy	Algebra I, Algebra II, Geometry, Integrated Math I, Integrated Math II, and Integrated Math III
Literacy and Numeracy	Algebra I, Algebra II, English I, English II, Geometry, Integrated Math I, Integrated Math II, and Integrated Math III
Science	Biology
Social Studies	U.S. History

5.2.7.5 TCAP (Grades 4-8)/EOC/Early Grades (Grade 3)

Composite Type	Subjects
Overall	Algebra I, Algebra II, Biology, English I, English II, Geometry, Integrated Math I, Integrated Math II, Integrated Math III, U.S. History, TCAP Math and English Language Arts (3–8), Science (5–8), and Social Studies (6–8)
Literacy	English I, English II, and TCAP English Language Arts (3–8)
Numeracy	Algebra I, Algebra II, Geometry, Integrated Math I, Integrated Math II, Integrated Math III, and TCAP Math (3–8)
Literacy and Numeracy	Algebra I, Algebra II, English I, English II, Geometry, Integrated Math I, Integrated Math II, Integrated Math III, and TCAP Math and English Language Arts (3–8)
Science	Biology and Science (5–8)
Social Studies	U.S. History and Social Studies (6–8)

5.2.7.6 CTE Students (Based on EOC)

Composite Type	Subjects
Overall	Algebra I, Algebra II, Biology, English I, English II, Geometry, Integrated Math I, Integrated Math II, Integrated Math III, and U.S. History
Literacy	English I, English II
Numeracy	Algebra I, Algebra II, Geometry, Integrated Math I, Integrated Math II, and Integrated Math III
Literacy and Numeracy	Algebra I, Algebra II, English I, English II, Geometry, Integrated Math I, Integrated Math II, and Integrated Math III
Science	Biology
Social Studies	U.S. History

5.2.7.7 CTE Concentrators (Based on EOC)

Composite Type	Subjects
Overall	Algebra I, Algebra II, Biology, English I, English II, Geometry, Integrated Math I, Integrated Math II, Integrated Math III, and U.S. History
Literacy	English I, English II
Numeracy	Algebra I, Algebra II, Geometry, Integrated Math I, Integrated Math II, and Integrated Math III
Literacy and Numeracy	Algebra I, Algebra II, English I, English II, Geometry, Integrated Math I, Integrated Math II, and Integrated Math III
Science	Biology
Social Studies	U.S. History

5.2.8 District and School Composites for Student Groups

As described in Sections 2.2.4.4 and 2.3.3, Tennessee uses value-added measures for student groups in their federal accountability system. For the student groups described in these sections, the gain and predictive growth measures are combined in the same way as the overall measure described in Sections 5.2.1 through 5.2.6.

These student group composites are available for districts and schools as single-year measures. The single-year composites are calculated in a process similar to the one detailed in Section 5.2.

District measures are available for the following grades in Math and English Language Arts:

- Grades 3–5 (for districts that administer Grade 2)
- Grades 4–5 (for all districts)
- Grades 6–8 (only includes TCAP subjects; does NOT include EOC subjects)
- Grades 9–12 (only includes EOC subjects)

School measures are available for the following grades in Math and English Language Arts:

- Composites with grade 3 measures (for schools that administer Grade 2):
- Composites without grade 3 measures

Depending on the eligible growth measures for the district and school, growth measures from the following assessments might be included.

Composite Type	Subjects
Math	TCAP Math (3–8), Algebra I, Algebra II, Integrated Math I, Integrated Math II, Integrated Math III, and Geometry
English Language Arts	TCAP English Language Arts (3–8), English I, and English II
Math and English Language Arts	TCAP Math and English Language Arts (3–8), Algebra I, Algebra II, English I, English II, Integrated Math I, Integrated Math II, Integrated Math III, and Geometry

6 Input Data Used in the Tennessee Growth Model

6.1 Assessment Data Used in Tennessee

For the analysis and reporting based on the 2023-24 school year, EVAAS receives the following assessments for use in the growth and/or projection models:

- TCAP Mathematics and English Language Arts and Science in grades 3–8
- TCAP Science in grades 3–8.
- TCAP Social Studies in grades 6–8
- EOC assessments in Algebra I, Algebra II, English I, English II, Biology, Geometry,
- Integrated Math I, Integrated Math II, Integrated Math III, and U.S. History
- ACT assessments in English, Math, Reading, and Science/Reasoning (based on "Junior Day" data)
- Advanced Placement (AP) assessments

Additionally, EVAAS receives the following assessments:

- TCAP Alt - Scores from this assessment are used only in student testing history.
- Dynamic Learning Maps (DLM) - Scores for this assessment are used only in student testing history. This assessment replaces the MSAA assessment.
- Grade 2 assessments in Mathematics and English Language Arts – Scores are provided for districts and/or schools that chose to administer these assessments. Scores from these assessments are only used as predictors to provide growth measures and projections for TCAP Math and ELA in grade 3.

Assessment files provide the following data for each student score:

- Administration
- District Number
- District Name
- School Number
- School Name
- Student Last Name
- Student First Name
- Student Middle Initial
- Student Date of Birth
- State Student ID Number
- Grade
- Test Grade

- Content Area Code
- Test Form
- Test Version
- Test Mode
- Tested/Attempted
- Student Not Tested
- RI Status
- School Type
- Test Status
- Student Total Raw Score - Points Earned
- Student Proficiency Classification
- Scoring Complete
- Student Scale Score

Some of this information, such as performance levels, is not relevant to the ACT.

6.2 Student Information

Student information is used in creating the web application to assist educators analyze the data to inform practice and assist all students with academic growth. This information might also be used to identify students whose data should be excluded from analysis. SAS receives this information in the form of various socioeconomic, demographic, and programmatic identifiers provided by TDOE. Currently, these categories are as follows:

- Gifted (Not Special Ed) (Y, N, U)
- Gender (M, F, U)
- Migrant Status (Y, N, U)
- English Learner (EL) (Y, N, U)
    - (No) No code
    - (Yes) Currently identified as English Learner or exited English as a Second Language program within the last 4 years
- Enrolled 50% or more
    - District level (Y, N)
    - School level (Y, N)
- ESL Transition
    - T1: 1 year since exiting ESL
    - T2: 2 years since exiting ESL
    - T3: 3 years since exiting ESL
    - T4: 4 years since exiting ESL

- Economically Disadvantaged (Y, N, U)
- Students with Disabilities (Y, N, U)
- Functionally Delayed (Not Special Ed) (Y, N, U)
- Career Technical Student (High School tests only) (Y, N, U)
- Career Technical Concentrator (High School tests only) (Y, N, U)
- Race
  - American Indian or Alaska Native
  - Asian
  - Black or African American
  - Hispanic
  - Native Hawaiian or Other Pacific Islander
  - White

In addition to the student information provided by TDOE, SAS creates a Lowest 25% identifier for students who are performing in the lowest 25% for a specific subject and grade at the district and school level. The details of the identification process vary for the gain and predictive models and can be found in Sections 2.2.4.5 and 2.3.3.2, respectively.

6.3 Teacher Information

In order to provide accurate and verified student-teacher linkages in the teacher growth models, Tennessee educators are given the opportunity to complete roster verification. This process enables teachers to confirm their class rosters for students in a particular subject, grade, and year, and it captures scenarios where multiple teachers have instructional responsibility for students. Administrators also verify the linkages as an additional check. Roster verification, therefore, increases the reliability and accuracy of teacher-level analyses.

Roster verification is completed within the TVAAS web application. TDOE provides EVAAS with access points to the data used to pre-populate roster verification. The data received through the roster verification process includes the following categories:

- Teacher-Level Identification
  - Teacher Name
  - Teacher License Number
- Student Linking Information
  - Student Last Name
  - Student First Name
  - Student Middle Initial
  - Unique Student ID (USID)
- Subjects and Tests for All State TCAP Achievement and EOC Assessments

    - Semester included for EOC Testing
- District and School Information (Numbers)
- Eligibility flag (Eligible or Ineligible roster)
- Percent of Instructional Responsibility (Instructional Time)
- Attendance flag (Instructional Availability)
    - F – Full
    - P – Partial

7 Business Rules

7.1 Assessment Verification for Use in Growth Models

To be used appropriately in any growth models, the scales of these assessments must meet three criteria:

1. **There is sufficient stretch in the scales** to ensure progress can be measured for both low-achieving students as well as high-achieving students. A floor or ceiling in the scales could disadvantage educators serving either low-achieving or high-achieving students.
2. **The test is highly related to the academic standards** so that it is possible to measure progress with the assessment in that subject, grade, and year.
3. **The scales are sufficiently reliable from one year to the next.** This criterion typically is met when there are a sufficient number of items per subject, grade, and year. This will be monitored each subsequent year that the test is given.

These criteria are checked annually for each assessment prior to use in any growth model, and Tennessee's current implementation include many assessments, such as TCAP Achievement, EOC, and college and career readiness assessments. These criteria are explained in more detail below.

7.1.1 Stretch

Stretch indicates whether the scaling of the assessment permits student growth to be measured for both very low- or very high-achieving students. A test "ceiling" or "floor" inhibits the ability to assess students' growth for students who would have otherwise scored higher or lower than the test allowed. It is also important that there are enough test scores at the high or low end of achievement, so that measurable differences can be observed.

Stretch can be determined by the percentage of students who score near the minimum or the maximum level for each assessment. If a much larger percentage of students scored at the maximum in one grade than in the prior grade, then it might seem that these students had negative growth at the very top of the scale when it is likely due to the artificial ceiling of the assessment. Percentages for all Tennessee assessments are well below acceptable values, meaning that these assessments have adequate stretch to measure value-added even in situations where the group of students are very high or low achieving.

7.1.2 Relevance

Relevance indicates whether the test is sufficiently aligned with the curriculum. The requirement that tested material correlates with standards will be met if the assessments are designed to assess what students are expected to know and be able to do at each grade level.

More information can be found at the following link:
https://www.tn.gov/education/instruction/academic-standards.html.

7.1.3 Reliability

Reliability can be viewed in a few different ways for assessments. Psychometricians view reliability as the idea that a student would receive similar scores if the assessment was taken multiple times. This type of reliability is important for most any use of standardized assessments.

7.2 Pre-Analytic Processing

7.2.1 Missing Grade

In Tennessee, the grade used in the analyses and reporting is the tested grade, not the enrolled grade. If a grade is missing on a grade-level test record (meaning 2–8), then that record will be excluded from all analyses. The grade is required to include a student's score in the appropriate part of the models and to convert the student's score into the appropriate NCE in the gain-based model.

7.2.2 Duplicate (Same) Scores

If a student has a duplicate score for a particular subject and tested grade in a given testing period in a given school, then the extra score will be excluded from the analysis and reporting.

7.2.3 Students with Missing School or Homeschool

If a student has a score with a missing school for a particular subject and grade in a given testing period and there is no duplicate score, then the score will be excluded from the analysis and reporting. The analysis also excludes student scores that are associated with a school that is identified as a homeschool.

7.2.4 Students with Multiple (Different) Scores in the Same Testing Administration

If a student has multiple scores in the same period for a particular subject and grade and the test scores are not the same, then those scores will be excluded from the analysis. Prior to exclusion, the student's records are examined to determine whether a single record can be selected based on all data available for the student.

7.2.5 Students with Multiple Grade Levels in the Same Subject in the Same Year

A student should not have different tested grade levels in the same subject in the same year. If that is the case, then the student's records are checked to see whether the data for two separate students were inadvertently combined. If this is the case, then the student data are adjusted so that each unique student is associated with only the appropriate scores. If the scores appear to all be associated with a single unique student, then scores that appear inconsistent are excluded from the analysis.

7.2.6 Students with Records That Have Unexpected Grade Level Changes

If a student skips more than one grade level (for example, moves from sixth in 2023 to ninth in 2024) or is moved back by one grade or more (that is, moves from fourth in 2023 to third in 2024) in the same subject, then the student's records are examined to determine whether two separate students were inadvertently combined. If this is the case, then the student data is adjusted so that each unique student is associated with only the appropriate scores. These scores are removed from the analysis if it is the same student.

7.2.7 Students with Records at Multiple Schools in the Same Test Period

If a student is tested at two different schools in a given testing period, then the student's records are examined to determine whether two separate students were inadvertently combined. If this is the case, then the student data is adjusted so that each unique student is associated with only the appropriate scores. When students have valid scores at multiple schools in different subjects, all valid scores are used at the appropriate school.

7.2.8 Outliers

Student assessment scores are checked each year to determine whether they are outliers in context with all the 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 TCAP and EOC Math subjects are examined simultaneously, and any scores that appear inconsistent, given the other scores for the student, are flagged. Outlier identification for college readiness assessments use all available college readiness data alongside state assessments in the respective subject area (for example, Math subjects with TCAP and EOC tests might be used to identify outliers with SAT or ACT). Furthermore, Grade 2 data are used for outlier identification with grade 3 scores.

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, then that outlier will not be used in the analysis, but it will be displayed on the student testing history on the TVAAS web application.

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 will provide a t-value of each comparison. Using this t-value, the growth models 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 removing a very high-achieving score.

For low-end outliers, the rules are:

- The percentile of the score must be below 50.
- The t-value must be below -2.5 for TCAP and EOCs when determining the difference between the score in question and the weighted combination of reference scores (otherwise known as the comparison score). In other words, the score in question must be at least 2.5 standard deviations below the comparison score. For other assessments, the t-value must be below -4.0.
- 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 4.5 for TCAP and EOCs when determining the difference between the score in question and the reference group of scores. In other words, the score in question must be at least 4.5 standard deviations above the comparison score. For ACT, the t-value must be above 5.0.
- The percentile of the comparison score must be below a certain value. This value depends on the position of the individual score in question but will need to be at least 30 to 50 percentiles below the individual percentile score.
- There must be at least three scores in the comparison score average.

7.2.9 Linking Records over Time

Each year, EVAAS receives data files that include student assessment data and file formats. These data are checked each year prior to incorporation into a longitudinal database that links students over time. Student test data and demographic data are checked for consistency year to year to ensure that the appropriate data are assigned to each student. Student records are matched over time using all data provided by the state, and teacher records are matched over time using the Unique ID and teacher's name.

7.3 Growth Models

7.3.1 Students Included in the Analysis

As described in Pre-Analytic Processing, student scores might be excluded due to the business rules, such as outlier scores.

For the gain, predictive, and projection models, the following students/test scores are excluded in accordance with TDOE policy:

- Students from home schools
- Students who "tested out of system"
- Students who do not have an overall SNT value of zero or missing
- Students who have a raw score equal to zero
- Test scores where "Attemptedness" is flagged as "No"
- Test scores that do not have an "Overall RI Status" of zero

In addition, students who are flagged as EL Recently Arrived Year 1 are excluded from the gain and predictive models for all assessments except for ACT. These students are included in projections and their scores will be included in future years as predictors.

For the gain model, all students are included in these analyses if they have assessment scores that can be used. The gain model uses all available TCAP Math and English Language Arts results for each student. Because this model follows students from one grade to the next and measures growth as the change in achievement from one grade to the next, the gain model assumes typical grade patterns for students. Students with non-traditional patterns, such as those who have been retained in a grade or skipped a grade, are treated as separate students in the model. In other words, these students are still included in the gain model, but the students are treated as separate students in different cohorts when these non-traditional patterns occur. This process occurs separately by subject since some students can be accelerated in one subject and not in another.

Specific to the teacher gain model, if a student does not have any prior test scores in the same subject in any prior year, then the student score is included in the model but will not be connected to any individual teachers.

For the predictive and projection models, a student must have at least three valid predictor scores (or two valid predictors for grade 3 and grade 4 projections) that can be used in the analysis. The predictors used to meet this minimum cannot be deemed outliers. (See Section 7.2.8 on Outliers.) These scores can be from any year, subject, and grade that are used in the analysis. In other words, the student's expected score can incorporate other subjects beyond the subject of the assessment being used to measure growth. The required three predictor scores are needed to sufficiently dampen the error of measurement in the tests to provide a reliable measure. If a student does not meet the three-score minimum, then that student is excluded from the analyses. It is important to note that not all students have to have the same three prior test scores; they only have to have some subset of three that were used in the analysis. Unlike the gain model, students with non-traditional grade patterns are included in the predictive model as one student. Since the predictive model does not determine growth based on consecutive grade movement on tests, students do not need to stay in one cohort from one year to the next. That said, if a student is retained and retakes the same test, then that prior score on the same test will not be used as a predictor for the same test as a response in the predictive model. This is mainly because very few students used in the models have a prior score on the same test that could be used as a predictor. In fact, in the predictive model, it is typically the case that a prior test is only considered a possible predictor when at least 50% of the students used in that model have those prior test scores.

For the teacher analysis in both the gain and predictive models, students are excluded from the teacher's analysis if they have a P value (or X in prior years) entered for instructional availability in the student-teacher linkages data.

The district and school models exclude students who are not enrolled 50% of the time within that district or school, in the current year. Additionally, in 2024, students from districts with low district-wide participation in the Integrated Math pathway were excluded from the district-level models for those subjects.

7.3.2 Minimum Number of Students to Receive a Report

The growth models require a minimum number of students in the analysis for districts, schools, and teachers to receive a growth report. This is to ensure reliable results.

7.3.2.1 District and School Model

For the gain model, the minimum student count to report an estimated average NCE *score* (that is, either entering or exiting achievement) is six students in a specific subject, grade, and year. To report an estimated NCE *gain* in a specific subject, grade, and year, there are additional requirements:

- There must be at least six students who are associated with the school or district in the subject, grade, and year and have scores for the subject, grade, and year of analysis.
- When considering all the students with valid Math and ELA scores in that school or district, in the grade and year of analysis, at least six students must have a prior year score in a subject for that subject to receive an estimated gain.
- At least one student in the school or district must have a "simple gain," which requires a valid test score in the current year and grade as well as the prior year and grade in the same subject. However, given the requirements outlined in the previous bullets, it is typically the case that at least six students have simple gains.
- For any district or school growth measures based on specific student groups, the same requirements described above apply for the students in that specific student group.

For example, to report an estimated NCE gain for school A in TCAP Math grade 5 for this year, there must be the following requirements:

- There must be at least six fifth-grade students with a TCAP Math grade 5 score at school A for this year.
- Of the fifth-grade students at school A this year across all subjects used in the gain model (Math and ELA*)*, there must be at least six students with a TCAP Math grade 4 score from the prior year.
- At least one of the fifth-grade students at school A this year must have a TCAP Math grade 5 score from this year *and* a TCAP Math grade 4 score from last year to allow for the calculation of a simple gain.

For the predictive model, the minimum student count to receive a growth measure is 10 students in a specific subject, grade, and year. These students must have the required three prior test scores needed to receive an expected score in that subject, grade, and year.

7.3.2.2 Teacher Model

The teacher gain model includes teachers who are linked to at least six students with a valid test score in the same subject, grade, and year. This requirement does not consider the percentage of instructional time that the teacher spends with each student in a specific subject and grade.

The teacher predictive model includes teachers who are linked to at least 10 students with a valid test score in the same subject/grade or course within a year. This requirement does not consider the percentage of instructional time the teacher spends with each student in a specific subject and grade.

For both the gain and predictive models, to receive a teacher report in a particular year, subject, and grade, there is an additional requirement. A teacher must have at least six Full Time Equivalent (FTE) students in a specific subject, grade, and year. 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 the teacher would not receive a teacher growth report. If another teacher taught 12 students for 50% of their instructional time, then that teacher would have six FTE students and would receive a teacher report. The instructional time attribution is obtained from the linkage roster verification process that is used in Tennessee.

The teacher gain model has an additional requirement. The teacher must be linked to at least five students, and one of these five students must have a "gain," meaning the same subject prior test score must come from the immediate prior year and prior grade. Students are linked to a teacher based on the subject area taught and the assessment taken. Students who have no prior testing data in the same subject area are not linked to the teacher for the analysis. Note that if a student repeats a grade, then the prior test data would not apply as the student has started a new cohort.

7.3.2.3 Student Groups

For any district or school growth measures based on specific student groups, the same requirements described above apply for the students in that specific student group. Note that student group reporting requires six students in the gain model and 10 students in the predictive model to be included in the TVAAS web reporting.

7.3.3 Student-Teacher Linkages

Student-teacher linkages are connected to assessment data based on the subject and identification information described in Section 6.3. The model will make adjustments to linkages if a student is claimed by teachers at a total percentage higher than 100% in an individual year, subject, and grade. If over-claiming happens, then the individual teacher's weight is divided by the total sum of all weights to redistribute the attribution of the student's test scores across teachers. Underclaimed linkages for students are not adjusted because a student can be claimed less than 100% for various reasons (such as a student who lives out of state for part of the year).