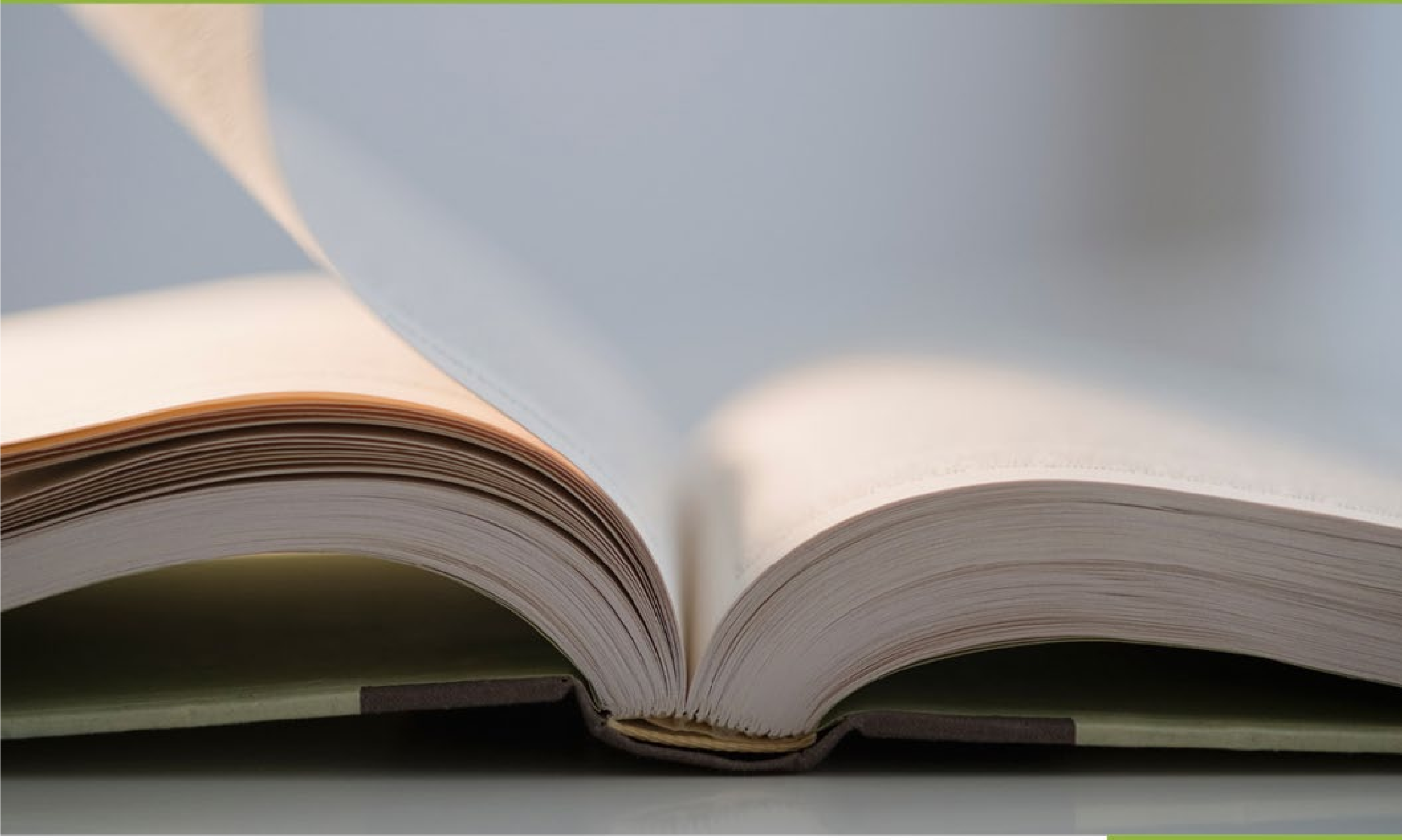
SAS® EVAAS

PVAAS Statistical Models and Business Rules

Prepared for Pennsylvania Department of Education



Contents

1 Introduction to Pennsylvania's Value-Added Reporting **1**

2 Statistical Models **3**

2.1 Overview of Statistical Models 3
2.2 Growth Standard Methodology 4
2.2.1 Overview 4
2.2.2 Why the Growth Standard Methodology is Needed 5
2.2.3 Common Scale in the Growth Standard Methodology 7
2.2.4 Technical Description of the Growth Standard Methodology 9
2.3 Predictive Methodology 16
2.3.1 Overview 16
2.3.2 Conceptual Explanation 17
2.3.3 Technical Description of the LEA/District, School, and Teacher Models 19
2.4 Projection Model 22
2.4.1 Overview 22
2.4.2 Technical Description 22
2.5 Outputs from the Models 23
2.5.1 Growth Standard Methodology 23
2.5.2 Predictive Methodology 25
2.5.3 Projection Model 27

3 Expected Growth **28**

3.1 Overview 28
3.2 Technical Description 28
3.3 Illustrated Example 28

4 Classifying Growth into Categories **31**

4.1 Overview 31
4.2 Use Standard Errors Derived from the Models 31
4.3 Define Growth Color Indicators in Terms of Standard Errors 31
4.4 Illustrated Examples of Categories 32
4.5 Interpret Growth Measures in Terms of Effect Size 34
4.6 Rounding and Truncating Rules 35

5 Composite Growth Measures **36**

5.1 Teacher Composites 36
5.1.1 Overview 36
5.1.2 Calculating the Index 37
5.1.3 Combining the Index Values Across Subjects, Grades, and/or Years 37
5.2 School AGI Composite Calculation (Three-Year Growth Measure in a Single Subject) 38
5.2.1 Overview of Three-Year Growth Measure for AGI 38
5.2.2 Sample Calculation for a School 38
5.2.3 Calculate Growth Standard Methodology Gain Across Grades 39
5.2.4 Calculate Growth Standard Methodology Standard Error Across Grades 39
5.2.5 Calculate Growth Standard Methodology-Based Composite Index Across Subjects 40
5.3 Act 13 Building Level Score and Future Ready PA Index Composite Calculation 40

5.3.1 Policy Decisions and Business Rules ... 40
5.3.2 Calculations ... 41
5.4 ESSA Composite Calculation for Designation and Exit Criteria ... 43
5.4.1 Policy Decisions and Business Rules ... 43
5.4.2 Calculations ... 43
5.4.3 Calculate Growth Standard Methodology Composite Index Across Subjects ... 44
5.4.4 Calculate Predictive-Methodology Index Across Subjects ... 45
5.4.5 Calculate the Combined Growth Standard and Predictive Methodology Composite Index Across Subjects ... 46
5.5 Available Composites ... 46

6 Input Data Used in the Pennsylvania Growth Model ... 48
6.1 Assessment Data Used in Pennsylvania ... 48
6.1.1 State Assessments ... 48
6.1.2 National Assessments ... 48
6.1.3 Locally Administered Assessments ... 48
6.2 Student Information ... 50
6.3 Teacher Information ... 51

7 Business Rules ... 53
7.1 Assessment Verification for Use in Growth Models ... 53
7.1.1 Stretch ... 53
7.1.2 Relevance ... 53
7.1.3 Reliability ... 53
7.2 Pre-Analytic Processing ... 54
7.2.1 Missing Grade ... 54
7.2.2 Duplicate (Same) Scores ... 54
7.2.3 Students with Missing LEAs/Districts or Schools for Some Scores but Not Others ... 54
7.2.4 Students with Multiple (Different) Scores in the Same Testing Administration ... 54
7.2.5 Students with Multiple Grade Levels in the Same Subject in the Same Year ... 55
7.2.6 Students with Records That Have Unexpected Grade Level Changes ... 55
7.2.7 Students with Records at Multiple Schools in the Same Test Period ... 55
7.2.8 Students Flagged as Homeschool ... 56
7.2.9 Students Flagged as Not Meeting Enrollment Criteria ... 56
7.2.10 Students Flagged as First-Year English Learners in the Current Year ... 56
7.2.11 Outliers ... 56
7.2.12 Linking Records over Time ... 57
7.3 Growth Models ... 57
7.3.1 Students Included in the Analysis ... 57
7.3.2 Minimum Number of Students to Receive a Report ... 60
7.4 Student-Teacher Linkages ... 61

1 Introduction to Pennsylvania's Value-Added Reporting

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 (LEA/district, school, student group, or teacher). For each of these models, there are business rules to ensure the growth measures reflect the policies and practices selected by the Commonwealth of Pennsylvania.

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. PVAAS growth models have been in use and available to educators in states since the early 1990s. These growth models were among the first in the nation to use sophisticated and robust statistical models that addressed these concerns.

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 Pennsylvania 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 Pennsylvania's value-added modeling based on the statistical models, business rules, policies, and practices selected by the Commonwealth of Pennsylvania and currently implemented by SAS® 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 LEA/district, school, student group, and teacher composites
- Input data
- Business rules

The Commonwealth of Pennsylvania has provided growth measures to Pennsylvania LEAs/districts, schools, and teachers since 2003. Known as PVAAS, the initial collaboration began with a pilot group of 100 LEAs/districts, and this expanded to statewide LEAs/districts and school value-added and student projection reporting by 2006. In 2014, Teacher Value-Added Reports also became available for the state. Over the years, reporting has expanded to include more assessments, such as student-level projections to PSAT, SAT, ACT, Advanced Placement, and ACCESS for ELLs assessments. In early 2022, LEA/district and school growth measures based on local assessments and CDTs became available for the first time, for participating LEAs/districts.

These reports are delivered through the PVAAS web application available at http://pvaas.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 focus 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 might 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 Pennsylvania, can account for these complexities and scenarios.

Pennsylvania'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 LEAs/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, PVAAS value-added measures typically have very low correlations with student characteristics. It is not necessary to make *direct* adjustments for student socioeconomic status or other student characteristics 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 modeling used for PVAAS growth measures:

> [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.
>
> Source: 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.

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 Pennsylvania does not make any direct adjustments for students' socioeconomic/demographic characteristics. Through this approach, the Pennsylvania Department of Education does not provide growth models to educators based on differential expectations for groups of students based on their backgrounds.

Value-added reporting is available based on Pennsylvania's state assessment program (PSSA and Keystones), Classroom Diagnostic Tests (CDTs), and other locally administered assessment data for LEAs/districts that submitted their data to PIMS.

Based on the available assessment data, there are two approaches to providing LEA/district, school, student group, and teacher growth measures.

- **Growth standard methodology (also known as the multivariate response model or MRM)** is used for tests given in consecutive grades, such as PSSA Math and English Language Arts in grades 3–8 to provide growth measures in grades 4–8. It is also used for tests given in consecutive time points, such as beginning-of-year (BOY) to end-of-year (EOY) in a specific grade and subject for non-CDT local assessments.
- **Predictive methodology (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 PSSA Science, Keystones, and CDT assessments.

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

- **The Projection model** is used for all assessments and provides a probability of obtaining a particular performance level/benchmark or higher on a given future assessment for individual students. This includes state assessments (PSSA and Keystones), PSAT, SAT, ACT, Advanced Placement (AP), and ACCESS for ELLs.

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

2.2 Growth Standard Methodology

2.2.1 Overview

The growth standard methodology measures growth between two points in time for a group of students; this is the case for tests given in consecutive grades such as PSSA Math and English Language Arts in grades 3–8 to provide growth measures in grades 4–8. This model is also used for non-CDT locally administered assessments. **More specifically, the growth standard methodology measures the change in relative achievement for a group of students based on the statewide achievement from one point in time to another.** Expected growth means that students maintained their relative achievement among the population of test-takers. More details are available in Section 3.

There are three separate analyses for PVAAS reporting based on the growth standard methodology: one each for LEAs/districts, schools, and teachers. Note that the student groups model leverages information from the school model. The LEA/district and school models are essentially the same; they perform well with the large numbers of students characteristic of LEAs/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 growth standard methodology 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 for use in fields with very large longitudinal data sets that tend to have missing data.

Value-added experts consider the growth standard methodology 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.

Note that the PSSA models are run separately from the non-CDT local assessment models with each model only using data from its assessment program. In other words, PSSA growth models use PSSA scores, and local assessment growth models only use scores from the same local assessment.

2.2.2 Why the Growth Standard Methodology 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 vs. 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 growth standard methodology uses the correlation between current and previous scores in the non-missing data to estimate means for all previous 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 growth standard methodology 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.[1] 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.

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 growth standard methodology calculates student growth.

[1] 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.

2.2.3 Common Scale in the Growth Standard Methodology

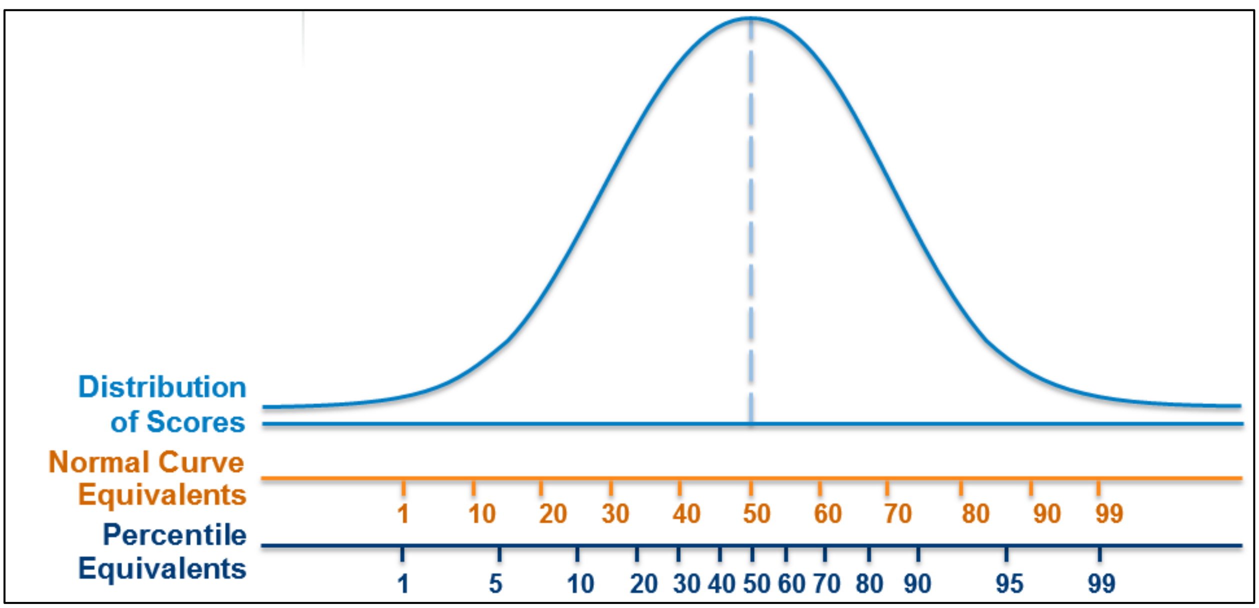
2.2.3.1 Why the Model Uses Normal Curve Equivalents

The growth standard methodology 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 as to 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, although percentile ranks are usually truncated below 1 and above 99, 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 Pennsylvania's state assessments, the average maximum NCE is approximately 130. While the growth standard methodology does not use truncated values, which could create an artificial floor or ceiling in students' test scores, the web reporting might show NCEs as integers from 1 to 99 for display purposes.

2.2.3.2 Sample Scenario: How to Calculate NCEs in the Growth Standard Methodology

The NCE distributions used in the growth standard methodology are based on a reference distribution of test scores in Pennsylvania. For state assessments, this reference distribution is the distribution of scores on a state-required test for all students in a given year. For non-CDT local assessments, the reference distribution is based on national norms provided by the assessment vendor. 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 growth standard methodology 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 test scores from Pennsylvania 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 1382 has a percentile rank of 43.3. The data show that 42.2% of students scored *below* 1382 while 44.4% of students scored *at or below* 1382. To calculate percentile ranks with discrete data, the usual convention is to consider half of the 2.2% reported in the Percent column to be "below" the cumulative percent and "half" above the cumulative percent. To calculate the percentile rank, half of 2.2% (1.1%) is added to 42.2% from Cumulative Percent to give you a percentile rank of 43.3, 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
1340	2,820	48,620	2.2	37.6	36.6	-0.344	42.76
1354	2,942	51,562	2.3	39.9	38.8	-0.285	44.00
1368	2,880	54,442	2.2	42.2	41.0	-0.226	45.23
1382	2,954	57,396	2.3	44.4	43.3	-0.169	46.45
1396	3,064	60,460	2.4	46.8	45.6	-0.110	47.69
1411	2,982	63,442	2.3	49.1	48.0	-0.051	48.93
1425	3,166	66,608	2.5	51.6	50.4	0.009	50.19

NCEs are obtained from the percentile ranks using the normal distribution. The table of the standard normal distribution (found in many textbooks[2]) or computer software (for example, a spreadsheet) provides the associated Z-score from a standard normal distribution for any given percentile rank. NCEs

[2] 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).

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 Growth Standard Methodology

2.2.4.1 Definition of the Linear Mixed Model

As a linear mixed model, the growth standard methodology for LEA/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.

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:[3]

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

[3] 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.

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.

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

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

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).

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

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

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 LEA/District and School Models

The LEA/district and school growth standard methodology 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 LEAs/districts, there is no need to include LEAs/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 a 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 growth standard methodology is able to use all available scores from each student.

Algebraically, the LEA/district growth standard 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} LEA/district. μ_{jkld} is the estimated mean score for this particular LEA/district, subject, grade, and year. ϵ_{ijkld} is the random deviation of the i^{th} student's score from the LEA/district mean.

The school model is represented as:

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

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

The growth standard methodology 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 LEA/district or school 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 (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 course of each 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 on 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 growth standard methodology 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 LEA/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 responsibility 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 LEA/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 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, and students are identified by the letter X-Z.

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 growth standard methodology provides LEA/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.

The list of student groups that receive growth measures is available in Section 2.5.

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

2.2.4.5 Accommodations to the Growth Standard Methodology for Missing 2019-20 Data Due to the Pandemic

This section describes accommodations to the gain model that were made for *2020-21* LEA/district and school reporting (2020-21 Teacher Reports were not calculated). However, this section does not apply to the 2021-22 and 2022-23 reporting since the immediate prior year is available to measure student growth.

In spring 2020, the COVID-19 pandemic required schools to close early and cancel statewide summative assessments. As a result, scores are not available for Pennsylvania's PSSA Math and ELA assessments based on the 2019-20 school year, and it is not possible to measure growth from the 2018-19 to the 2019-20 school years or from the 2019-20 to the 2020-21 school years. For the growth standard methodology based on PSSA Math and ELA, the 2020-21 reporting measures growth from the 2018-19 school year to the 2020-21 school year. LEA/district and school growth measures are provided for the 2020-21 reporting, but no teacher growth measures were calculated for that year.

From a technical perspective, the growth standard methodology for PSSA is essentially the same as it has been in previous years except that growth is measured over two years rather than one year. However, the interpretation of these growth measures changes slightly in two notable ways.

First, because the models provide two-year growth measures, the growth measure for grades where students transition from one school to another will then include growth from the feeder school(s) as well as the receiver school. For example, a middle school with grades 6–8 could receive a growth measure for sixth grade based on the students' growth in sixth grade as well as their growth from the feeder elementary school(s) in fifth grade.

In other words, it is not possible to parse out the individual contribution of the middle school in 2021 sixth grade apart from those from the elementary school(s) in fifth grade because of the missing year of test scores in 2020. For the LEA/district-level growth measures and for the non-transition grades, the two-year growth measures (from 2019 to 2021) are still solely representative of growth within the specific LEA/district and the non-transition grades for the school are still solely representative of growth within the specific school.

Second, at a particular school, the growth of certain groups of students are not represented in the two-year measures as they would be in two separate one-year growth measures. For example, it is not possible to measure the growth of grade 4 students in 2021 because there is no grade 3 data from last year (2020) and no statewide assessment to use from grade 2 in 2019. Similarly, it is not possible to report grade 8 growth from 2020 because there is no exiting achievement for these students in their last year at the school as eighth graders in 2020.

Despite these differences, the conceptual explanation of the 2020-21 growth measures is the same as it has always been: these growth measures compare students' exiting achievement with their entering achievement over two points in time.

2.3 Predictive Methodology

2.3.1 Overview

Tests that are not given in consecutive grades or tests where prior performance is used to predict performance on another require a different modeling approach from the growth standard methodology. The predictive model is used for such assessments in Pennsylvania. **The predictive methodology is a regression-based model where growth is a function of the difference between students' predicted scores with their actual scores.** Predicted growth is met when students with a LEA/district, school, or teacher made the same amount of growth as students with the average LEA/district, school, student group, or teacher.

Like the growth standard methodology, there are three separate analyses for PVAAS reporting based on the predictive methodology: one each for LEAs/districts, schools, and teachers. Note that the student groups model is based on the school model. The LEA/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 methodology is known as the univariate response model (URM), a linear mixed model and, more specifically, an analysis of covariance (ANCOVA) model.

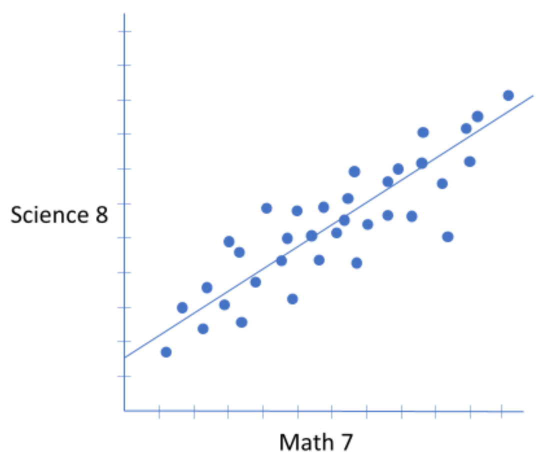
The key advantages of the predictive methodology can be summarized as follows:

- 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 the minimum required number of predictors of the response variable in any subject and grade.
- It allows educators to benefit from all tests, even when tests are on differing 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

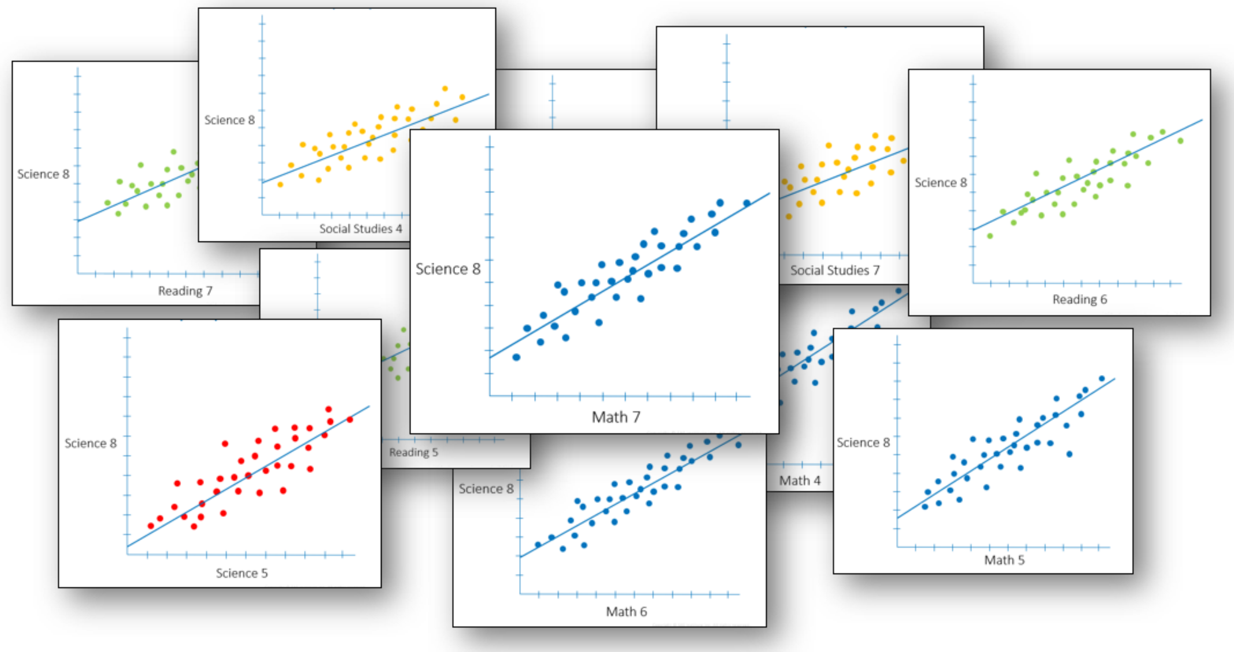
As mentioned above, the predictive methodology is ideal for assessments given in non-consecutive grades, such as PSSA Science and Keystones. Consider all students who tested in PSSA Science in grade 8 in a given year. The growth standard methodology is not possible since there isn't a Science test in the immediate prior grade. However, these students might have a number of prior test scores in PSSA Math and ELA in grades 3–8. These prior test scores have a relationship with PSSA Science in grade 8, meaning that how students performed on these tests can predict how the students perform on PSSA Science in grade 8. The growth model does not assume what the predictive relationship will be; instead, the actual relationships observed by the data define the relationships. This is shown in Figure 3 below where each dot represents a student's prior score on PSSA Math grade 7 plotted with their score on PSSA Science grade 8. The best-fit line indicates how students with a certain prior score on PSSA Math grade 7 tend to score, on average, on PSSA Science grade 8. This illustration is based on one prior test; the predictive methodology uses many prior test scores from different subjects and grades.

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



Some subjects and grades will have a greater relationship to PSSA Science in grade 8 than others; however, the other subjects and grades still have a predictive relationship. For example, prior Math scores might have a stronger predictive relationship to PSSA Science in grade 8 than prior ELA scores, but how a student performs on the PSSA ELA test typically provides an idea of how we might expect a student to perform on average on PSSA Science test. This is shown in Figure 4 below, where there are a number of different tests that have a predictive relationship with PSSA Science in grade 8. All of these relationships are considered together in the predictive model, with some tests weighted more heavily than others.

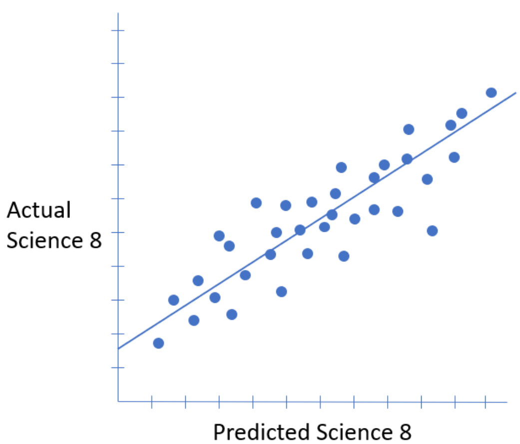
Figure 4: Relationships Observed in the Statewide Data Inform the Predictive Methodology



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 a predicted score based on their own prior testing history. In practical terms, the predicted 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 predicted and actual scores for a group of students.

Figure 5: Relationship Between Predicted Score and Actual Score for Selected Subject and Grade



The predicted scores can be aggregated to a specific LEA/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 predicted score) for a group of students. Unlike the growth standard methodology, the actual score and predicted score are reported in the scaling units of the test rather than NCEs.

Note that the Keystone assessments do not use any CDT data as prior test scores to calculate a predicted score for a student. In contrast, CDT assessments use prior CDT assessments as well as prior PSSA and Keystones assessments as predictors.

2.3.3 Technical Description of the LEA/District, School, and Teacher Models

The predictive methodology has similar approaches for LEAs/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 LEA/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.

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

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 growth standard methodology 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} LEA/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, β terms, and sometimes α_j). The parameter estimates (denoted with "hats," e.g., $\hat{\mu}$, $\hat{\beta}$) are obtained using all students that have an observed value for the specific response and have three predictor scores. The resulting prediction equation for the i^{th} student is as follows:

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

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

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

Note that *C* in equation (16) is not the same as *C* in equation (4). This matrix is estimated using the EM (expectation maximization) 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 (LEA/district, school, or teacher) covariance matrix. This is accomplished by providing scores to the EM algorithm that are centered around group means (i.e., 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 (or two predicators for PSSA Science in grade 4 and CDT in grades 3 and 4). Given such a matrix, the vector of estimated regression coefficients for the projection equation (15) can be obtained as:

$$\hat{\beta} = C_{xx}^{-1} c_{xy} \tag{17}$$

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" LEA/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 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.

$$\hat{y}_i = \hat{\mu}_y + \hat{\beta}_1(x_{i1} - \hat{\mu}_1) + \hat{\beta}_2(x_{i2} - \hat{\mu}_2) + \cdots \tag{18}$$

The $\hat{y}_i$ term is nothing more than a composite of all the student's past scores. It is a one-number summary of the student's level of achievement prior to the current year, and this term is called the predicted 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 Keystone Algebra I than when it is Literature, 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 LEA/district, school, or teacher.

The second step in the predictive methodology is to estimate the group effects (α_j) using the following ANCOVA model.

$$y_i = \gamma_0 + \gamma_1 \hat{y}_i + \alpha_j + \epsilon_i \tag{19}$$

In the predictive methodology, the effects (α_j) are considered random effects. Consequently, the $\hat{\alpha}_j$ terms are obtained by shrinkage estimation (empirical Bayes).[4] The regression coefficients for the ANCOVA model are given by the γ terms.

In the predictive model, there is an adjustment for Keystones Algebra I and CDT Algebra I that considers the enrolled grade of the student since 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 within a LEA/district, school, or teacher take Algebra I.

2.3.3.1 Predictors Used for Measuring Growth on Different Assessments

When measuring growth for…	These prior test scores are used to calculate the predicted score when available…
PSSA Science 4 and 8	PSSA Math and ELA 3-7; PSSA Science 4
Keystones Algebra I, Biology, and Literature	PSSA Math and ELA 3-8; PSSA Science 4 and 8; Keystone Algebra I (used for Keystone Biology and Literature growth); and Keystone Biology (used for Keystone Literature growth)
CDT assessments	All prior PSSA Math, Reading/ELA, and Science, and Keystones assessments and same year/subject BOY CDTs. For CDT 3rd grade, predictors include BOY same year Math/ELA CDTs as predictors

2.3.3.2 Accommodations to the Predictive Methodology for Missing 2019-20 Data Due to the Pandemic

This section describes whether any accommodations were necessary for the predictive model in the *2020-21* reporting. However, this section does not apply to the 2021-22 or 2022-23 reporting since the immediate prior year is available to measure student growth.

In spring 2020, the COVID-19 pandemic required schools to close early and cancel statewide summative assessments. As a result, statewide scores are not available for most of Pennsylvania's PSSA and Keystone exams based on the 2019-20 school year, and it is not possible to measure growth from assessment data through the 2019-20 school year. For the predictive methodology, the 2020-21

[4] 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).

reporting measures growth using students' predictors through the 2018-19 school year and then compares their predicted score to their actual performance on the 2020-21 assessment. LEA/district and school growth measures are provided for the 2020-21 reporting, but no teacher growth measures were calculated for that year.

As a reminder, the predictive methodology is used to measure growth for assessments given in non-consecutive grades, such as PSSA Science and Keystones, as well as CDTs in the Keystone content areas. Because these assessments are not administered every year in consecutive grades, *it has always been possible* that students do not have any test scores in the *immediate* prior year. The model can provide a robust estimate of students' entering achievement for the course by using all other available test scores from other subjects, grades, and years.

In other words, the predictive methodology does not require any technical adaptations to account for the missing year of data, and the interpretation of the results is similar to a typical year of reporting.

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 of achievement to future assessments. A projection is reported as a probability of obtaining a specific achievement score or above on an assessment, such as a 70% probability of scoring Proficient 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. Due to atypical schooling experiences during and after the pandemic, some projections are based on the cohort of students who took the assessment in the 2018-19 or 2021-22 school year rather than the 2022-23 school year as this represents a more typical schooling experience. Projections are available for state assessments as well as to college readiness assessments (ACT, PSAT, SAT, and Advanced Placement) and the ACCESS for ELLs assessment.

Student 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 academic 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 methodology 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.

$$y_i = \mu_y + \alpha_j + \beta_1(x_{i1} - \mu_1) + \beta_2(x_{i2} - \mu_2) + \cdots + \epsilon_i \qquad (20)$$

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

$$\hat{y}_i = \hat{\mu}_y \pm \hat{\alpha}_j + \hat{\beta}_1(x_{i1} - \hat{\mu}_1) + \hat{\beta}_2(x_{i2} - \hat{\mu}_2) + \cdots - \quad (21)$$

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 (16), (17), and (18) and will not be repeated here.

Once the parameter estimates for the projection equation have been obtained, projections can be made for any student with any set of predictor values. However, 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 two predictors in the case of PSSA ELA, Math, and Science for grade 4). 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 Proficient on a future end-of-grade 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.

$$Prob(\hat{y}_i \geq b) = \Phi\left(\frac{\hat{y}_i - b}{SE(\hat{y}_i)}\right) \quad (22)$$

2.5 Outputs from the Models

2.5.1 Growth Standard Methodology

For state assessments, the growth standard methodology is used for courses where students test in consecutive grade-given tests. As such, **the growth standard methodology uses PSSA in Math and ELA in grades 3–8 to provide LEA/district, school, and teacher growth measures in the following content areas:**

- PSSA Math in grades 4–8 (5–8 for 2020-21 reporting)
- PSSA ELA in grades 4–8 (5–8 for 2020-21 reporting)

No Teacher Reports are available based on the 2020-21 school year due to the missing assessment data in the prior school year.

The growth standard methodology also provides LEA/district-level and school-level growth measures only in the following content areas for LEAs/districts that submitted their locally administered assessment data:

- Acadience (DIBELS Next) ELA Composite in grades K-6
- Acadience (DIBELS Next) ELA Oral Reading, Fluency in grades 2-6
- Acadience (DIBELS Next) Math Composite in grades K-6
- Aimsweb ELA Composite in grades K and 2-8
- Aimsweb ELA Oral Reading Fluency in grades 1-8
- Aimsweb Math Composite in grades K-8
- DIBELS ELA Oral Reading Fluency in grades 1-8
- DIBELS ELA Composite in grades K-8
- EasyCBM Math in grades K-8
- Exact Path ELA in grades 5-8
- Exact Path Math in grades 2-8
- Exact Path Reading in grades 2-8
- FastBridge Math in grades K-8
- i-Ready Math Overall in grades K-8
- i-Ready Reading Overall in grades K-8
- MAP Math in grades K-12
- MAP Reading in grades K-12
- STAR Math in grades 1-12
- STAR Reading in grades K-12

This list is based on the current year of reporting. Exact test, subject, grades from previous years can vary.

In addition to the mean scores and mean gain for an individual subject, grade, and year, the growth standard methodology 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; not available on the web for 2022-23 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 LEA/district and school models and in equations (9) and (10) for the teacher model. More details about LEA/district, school, and teacher composites across subjects, grades, and years are available in Section 5.

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

- LEA/District growth measures
  - Overall students (for state assessments and local assessments)
  - Student Groups (for state and local assessments)
    - American Indian/Alaskan Native
    - Asian
    - Black
    - Combined Ethnicity (state assessments only)
    - Economically Disadvantaged
    - English Learner
    - Foster (state assessments only)
    - Hispanic
    - Homeless (state assessments only)
    - Lowest performing 33% of students
    - Military-connected family (state assessments only)
    - Two or More Races
    - Hawaiian/Pacific Islander
    - Students with IEPs
    - Students with GIEPs (provided from PIMS)
    - White
- School growth measures
  - Overall students (for state assessments and local assessments)
  - Student Groups (for state and local assessments)
    - American Indian/Alaskan Native
    - Asian
    - Black
    - Combined Ethnicity (state assessments only)
    - Economically Disadvantaged
    - English Learner
    - Foster (state assessments only)
    - Hawaiian/Pacific Islander
    - Hispanic
    - Homeless (state assessments only)
    - Lowest performing 33% of students
    - Military-connected family (state assessments only)
    - Two or More Races
    - Students with IEPs
    - Students with GIEPs (provided from PIMS)
    - White
- Teacher growth measures
  - Overall students (for state assessments only)

2.5.2 Predictive Methodology

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

- PSSA Science in grades 4 and 8 (only grade 8 available for 2020-21 reporting)
- Keystone Exams in Algebra I, Biology and Literature

The predictive methodology also provides LEA/district- and school-level growth measures only in the following content areas:

- Classroom Diagnostic Test (CDTs) in CDT Math, ELA, and Science in grades 3-8 and Algebra I, Biology, and English Literature

In addition to the mean scores and growth measures for an individual subject, grade, and year, the predictive methodology can also provide multi-year average growth measures (up to three years) for each subject and grade or course. Note that multi-year average growth measures are not available on the web for the 2022-23 reporting.

Collectively, the different models provide metrics for a variety of purposes within the Commonwealth of Pennsylvania. They are summarized in the table below:

- LEA/District growth measures
  - Overall students (for state assessments and CDTs)
  - Student Groups (for state assessments and CDTs)
    - American Indian/Alaskan Native
    - Asian
    - Black
    - Combined Ethnicity (state assessments only)
    - Economically Disadvantaged
    - English Learner
    - Foster (state assessments only)
    - Hawaiian/Pacific Islander
    - Hispanic
    - Homeless (state assessments only)
    - Lowest performing 33% of students
    - Military Family (state assessments only)
    - Two or More Races
    - Students with IEPs
    - Students with GIEPs (provided from PIMS)
    - White
- School growth measures
  - Overall students (for state assessments and CDTs)
  - Student Groups (for state assessments and CDTs)
    - American Indian/Alaskan Native
    - Asian
    - Black
    - Combined Ethnicity (state assessments only)
    - Economically Disadvantaged
    - English Learner
    - Foster (state assessments only)
    - Hawaiian/Pacific Islander

    - Hispanic
    - Homeless (state assessments only)
    - Lowest performing 33% of students
    - Military Family (state assessments only)
    - Two or More Races
    - Students with IEPs
    - Students with GIEPs (provided from PIMS)
    - White
- Teacher growth measures
    - Overall students (for state assessments only)

Note that more details about LEA/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, college readiness assessments (ACT, PSAT, SAT, Advanced Placement), and the ACCESS for ELLs assessment.

- PSSA Math and ELA in grades 4–8
- PSSA Science in grades 4 and 8
- Keystone Algebra I, Biology, and Literature
- PSAT 8/9 in Mathematics and Reading and Writing
- PSAT NMSQT Mathematics and Reading and Writing
- SAT Mathematics and Reading and Writing
- ACT English, Math, Reading, and Science
- AP Biology, Calculus AB, English Language and Composition, English Literature Composition, Psychology, Statistics, United States Government and Politics, and United States History
- ACCESS for ELLs in grades 1–12

More specifically, PSSA projections are typically provided one or two grade levels above a student's last tested grade, such as projections to grades 6 and 7 for students who most recently tested in grade 5.

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 are useful for the interpretation and application of growth measures. Note that these explanations were essentially the same for the 2020-21 reporting as they are in typical years; it is just the time period and available prior test scores that changed for the 2020-21 reporting due to the missing year of data from the 2019-20 school year.

3.2 Technical Description

Both the growth standard and predictive methodologies 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 methodologies 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 growth standard methodology, expected growth means that students maintained the same relative position with respect to the statewide student achievement that year. In the predictive methodology, expected growth means that students with a LEA/district, school, or teacher made the same amount of growth as students with the average LEA/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 LEA/district (or school or teacher) estimates above zero and approximately half of the LEA/district (or school or teacher) estimates below zero.

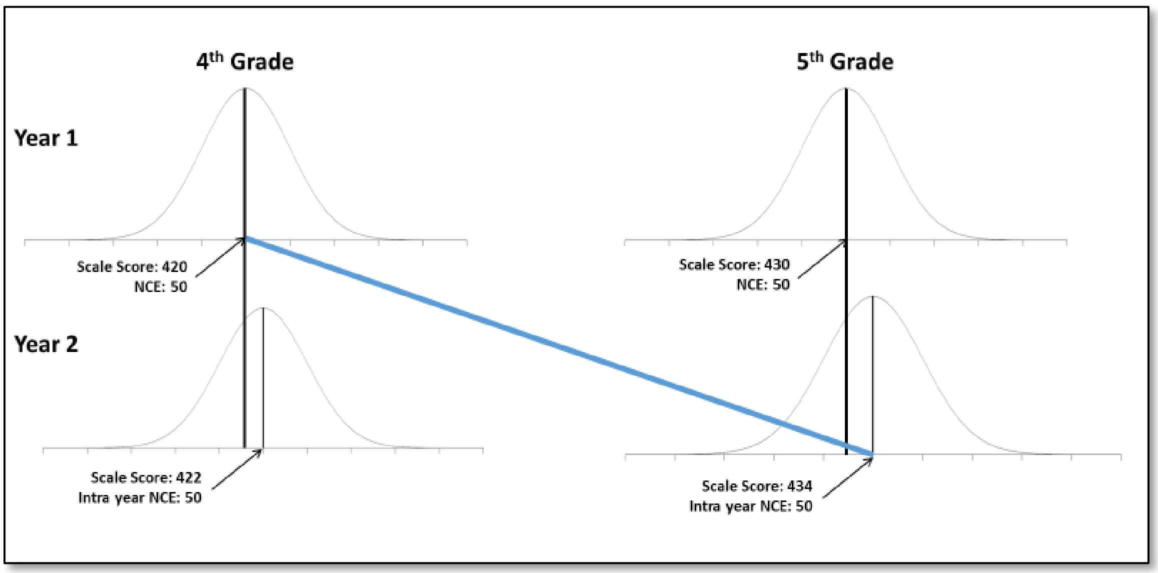
A change in assessments or scales from one year to the next does not present challenges to calculating expected growth. Through the use of NCEs, the growth standard methodology converts any scale to a relative position, and the predictive model already uses prior test scores from different scales to calculate the predicted 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 growth standard methodology 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 50^{th} NCE (similar to the 50^{th} 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.

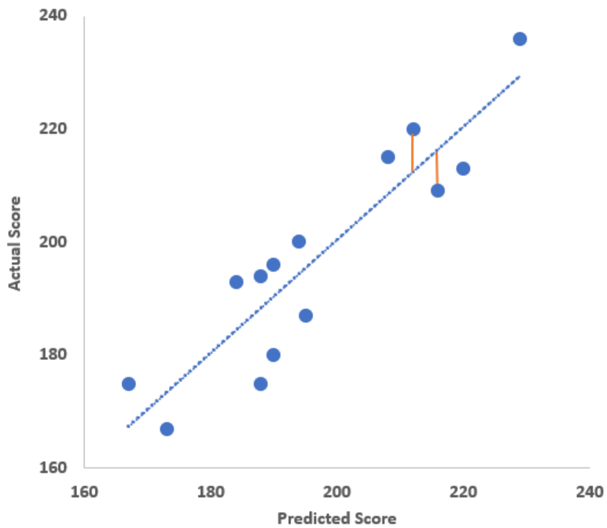
Note that the actual gain calculations are much more robust than what is presented here; as described in the previous section, the models can address students with missing data, team teaching, and all available testing history.

Figure 6: Intra-Year Approach Example for the Growth Standard Methodology



In contrast, in the predictive methodology, 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 methodology. The graph plots each student's actual score with their predicted score. Each dot represents a student, and a best-fit line will minimize the difference between all students' actual and predicted scores. Collectively, the best-fit line indicates what expected growth is for each student – given the student's predicted 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 predicted 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 (LEA/district, school, or teacher).

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



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. Pennsylvania's growth model has five categories for LEAs/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 color 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 LEA/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. The teacher model also takes into account shared instruction, such as 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, 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 amount of evidence that students in a LEA/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 LEAs/districts, schools, and teachers to incorporate standard errors to the value-added reporting.

4.3 Define Growth Color 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 PVAAS 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 color 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

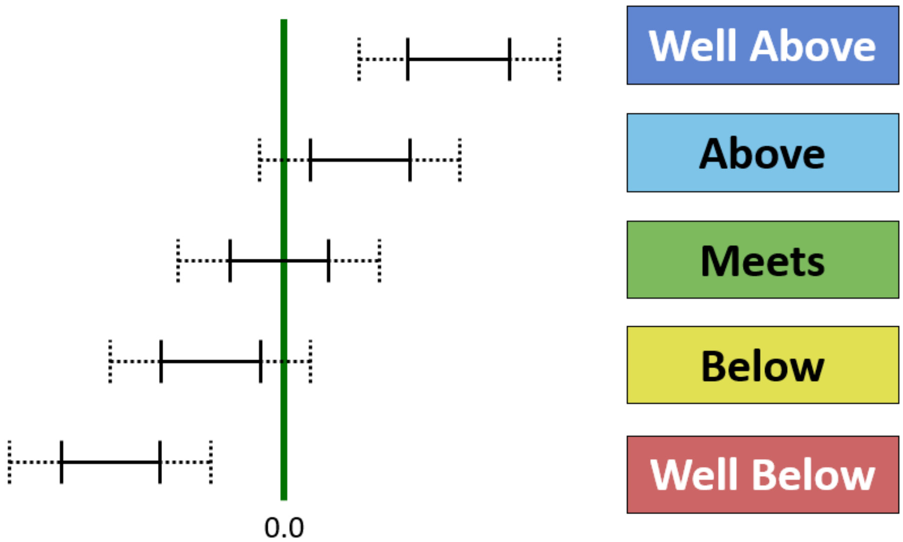
There are two ways to visualize how the growth measure and standard error relate to expected growth and how these can be used to create categories.

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

- **Well Above** indicates that the growth measure is two standard errors or more above expected growth (0). This is significant evidence of exceeding the growth standard.
- **Above** indicates that the growth measure is at least one but less than two standard errors above expected growth (0). This is moderate evidence of meeting the growth standard.
- **Meets** indicates that the growth measure is less than one standard error above expected growth (0) and no more than one standard error below expected growth (0). This is evidence of meeting the growth standard.
- **Below** indicates that the growth measure is more than one but no more than two standard errors below expected growth (0). This is moderate evidence of not meeting the growth standard.
- **Well Below** is an indication that the growth measure is less than or equal to two standard errors below expected growth (0). This is significant evidence of not meeting the growth standard.

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 Well Above 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 Well Below category. Above and Below 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. Meets 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 frame the categories is to create a growth index, which is calculated as shown below:

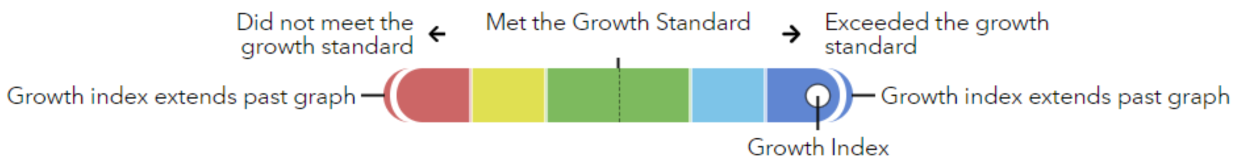
$$Growth\ Index = \frac{Growth\ Measure - Expected\ Growth}{Standard\ Error\ of\ the\ Growth\ Measure} \tag{23}$$

The growth index is similar in concept to a Z-score or t-value, and it communicates as a single metric the amount of 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:

- **Well Above** indicates significant evidence that students exceeded the growth standard. The growth index is 2 or greater.
- **Above** indicates moderate evidence that students exceeded the growth standard. The growth index is between 1 and 2.
- **Meets** indicates evidence that students met the growth standard. The growth index is between -1 and 1.
- **Below** indicates moderate evidence that students did not meet the growth standard. The growth index is between -2 and -1.
- **Well Below** indicates significant evidence that students did not meet the growth standard. The growth index is less than -2.

This is represented in the growth color indicator bar in Figure 9, which is similar to what is provided in the LEA/District and School Value-Added Reports in the PVAAS 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 Color 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 Interpret Growth Measures in Terms of Effect Size

Effect size is another common statistical metric that converts growth measures into a common scale across different models and assessments for easier comparison and also indicates the growth measure's magnitude or practical significance. Effect sizes are available in the LEA/District and School Value-Added Reports for PSSA, Keystones, and CDT assessments.

The effect size is calculated as the growth measure divided by the student-level standard deviation of growth. For the growth standard methodology, students typically have a current and a prior year NCE, which is used to derive a student-level gain. For the predictive methodology, student-level gain is defined as the student's observed minus predicted score.

The standard deviation of the student-level distribution of growth is calculated for each year, subject, and grade. Dividing the growth measures by the standard deviation provides the "effect size," and it indicates the practical significance regarding the group of students and whether they met, exceeded, or fell short of expected growth.

The effect size can be classified as small, medium, or large to assist with interpretation and whether any differences in student performance are meaningful. Various researchers have offered thoughts on what defines a small, medium, and large effect size.

- Cohen describes 0.20 as small, 0.50 as medium, and 0.80 as large (Cohen, Jacob. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Mahwah, NJ: Lawrence Erlbaum, 1988).
- Hattie describes an effect size of 0.40 as the average seen across all interventions, and 0.40 as the "hinge point" (Hattie, John, *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement.* London: Routledge, 2008).
- Kraft suggested < 0.05 as small, 0.05 to 0.20 as medium, and > 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).

All of the 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 growth color indicator 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 sections below describe the calculation of composites for teachers first, then LEAs/districts, and finally schools. Because teacher-specific reporting was not available for the 2019-20 and 2020-21 school year and LEA/district and school reporting were not available for the 2019-20 school year, the composites for the 2022-23 differ somewhat from what has been calculated historically.

5.1 Teacher Composites

5.1.1 Overview

The key policy decisions for 2022-23 teacher composites are summarized as follows:

- For teachers who have two consecutive years of teacher-specific reporting available (2021-22 and 2022-23 school years), a two-year trend composite is calculated using all subjects and grades.
- For teachers who have only 2022-23 teacher-specific reporting available and do not have 2021-22 teacher-specific reporting available, a single-year trend composite is calculated using all subjects and grades from the 2022-23 school year.
- The composite weights each subject/grade/year and course/year equally.
- The composite includes PSSA Math, ELA, Science, and any Keystone assessments.

Note that, in previous years, a multi-year composite was calculated each year but was not used in the Pennsylvania teacher evaluation until it contains three consecutive school years of value-added data.

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

1. Calculate a growth index for each assessment in the applicable subject, grades, and/or years.
2. Calculate a composite index that combines all applicable growth indices from the individual assessments where each growth index for each assessment is weighed equally.
3. This calculation uses non-rounded values, and then applies rounding or truncating rules as described in Section 4.6.

If a teacher does not have value-added measures from both the growth standard and predictive methodologies, then the composite index would be based on the model for which the teacher does have reporting.

The following sections illustrate this process using value-added measures from a sample teacher, which are provided in Table 3.

Table 3: Sample Teacher Value-Added Information

Year	Subject	Grade	Growth Measure	Standard Error	Average Growth Index
1	Science	8	15.20	7.00	2.17
1	Math	7	3.50	1.50	2.33
2	ELA	8	0.50	1.40	0.36
2	Math	8	4.50	1.60	2.81

5.1.2 Calculating the Index

The teacher in the example above has taught a mixture of subjects and grades in Years 1 and 2. All these measures are used in the two-year composite calculation. As explained in earlier sections, the model produces a value-added measure and standard error for each year, subject, and 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.

In the example above, the value-added measures for Math and ELA are on the NCE scale, whereas the value-added measure is reported in the scale score units in Science. An index is calculated for each of these measures by dividing the value-added measure by its standard error and is given in the final column.

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. This standardized statistic has a standard error of 1.

5.1.3 Combining the Index Values Across Subjects, Grades, and/or Years

To calculate the overall composite that uses value-added information for all available subjects and grades in Years 1 and 2, the first step is to average the index values. In the above example, this would look like the following using the numbers from the last column of Table 3:

$$Avg.Index = \left(\frac{1}{4}\right)(2.17) + \left(\frac{1}{4}\right)(2.33) + \left(\frac{1}{4}\right)(0.36) + \left(\frac{1}{4}\right)(2.81) = 1.92 \quad (24)$$

Note that this example reports rounded values for display purposes, but the actual calculation uses non-rounded values until the final index.

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 uses a standard statistical practice to ensure that the final index has a standard error of 1. This correction is simple, but to derive where it comes from, the standard error of an average index can be found using the following formula.

$$Standard\ Error\ Avg.Index = \frac{1}{4}\sqrt{1^2 + 1^2 + 1^2 + 1^2} = \frac{\sqrt{4}}{4} = \frac{1}{\sqrt{4}} = 0.50 \quad (25)$$

To calculate the new index, the average of the index values would be divided by the new standard error of the average index. Therefore, to get the new index value the average of the indexes is multiplied by square root of the number of measures that went into it.

$$Composite\ Index = \frac{1.92}{\left(1/\sqrt{4}\right)} = 1.92\sqrt{4} = 3.84 \tag{26}$$

5.2 School AGI Composite Calculation (Three-Year Growth Measure in a Single Subject)

5.2.1 Overview of Three-Year Growth Measure for AGI

The section describes how three-year growth measures are typically calculated for the AGI, but this measure is not available for the 2022-23 reporting.

The school three-year growth measure is calculated by averaging all the grade-specific growth measures that exist across all three years for that subject. It is a composite across years, not subjects. Since each measure is given equal weight, the three-year AGI might not be the same as the average of the growth measures for individual year AGIs if a school has changed grade configurations over the past three years. Note: In Math and ELA, students can be included in more than one growth measure across years, such as the 2023 PSSA ELA grade 5 measure as well as the 2022 PSSA ELA grade 4 measure. As a result, these growth measures cannot be assumed to be independent, and the covariance term is calculated during the modeling process. These concepts are explained in more detail below.

For measures that are independent (the Keystones), the covariance term is zero.

The three-year AGI is the three-year growth measure divided by the three-year standard error, to which rounding rules are then applied.

5.2.2 Sample Calculation for a School

The following sections illustrate a sample calculation using value-added measures from a sample elementary school in Math, assuming it has results for grades 4 and 5 across three years.

Table 4: Sample School Value-Added Information

Year	Subject	Grade	Growth Measure	Standard Error
1	Math	4	3.30	0.70
1	Math	5	-1.10	1.00
2	Math	4	2.00	0.50
2	Math	5	2.40	1.10
3	Math	4	-0.30	0.60
3	Math	5	3.80	0.70

5.2.3 Calculate Growth Standard Methodology Gain Across Grades

The school composite gain is a simple average of the six gain-based value-added measures. More specifically, the gain is calculated using the following formula:

$$Composite\ Gain = \tfrac{1}{6}Math_{4_{Year\ 1}} + \tfrac{1}{6}Math_{5_{Year\ 1}} + \tfrac{1}{6}Math_{4_{Year\ 2}} + \tfrac{1}{6}Math_{5_{Year\ 2}} + \tfrac{1}{6}Math_{4_{Year\ 3}} + \tfrac{1}{6}Math_{5_{Year\ 3}}$$
$$= \left(\tfrac{1}{6}\right)(3.30) + \left(\tfrac{1}{6}\right)(-1.10) + \left(\tfrac{1}{6}\right)(2.00) + \left(\tfrac{1}{6}\right)(2.40) + \left(\tfrac{1}{6}\right)(-0.30) + \left(\tfrac{1}{6}\right)(3.80) = 1.68 \qquad (27)$$

Note that this example reports rounded values for display purposes. The actual calculation uses non-rounded values until the final index.

5.2.4 Calculate Growth Standard Methodology Standard Error Across Grades

As discussed above, it cannot be assumed that the gains in the composite are independent because it is likely some of the same students are represented in different value-added gains. If this was only a three-year average for the same subject and grade, then the gains could be considered independent since those would represent different students each year. Again, to demonstrate the impact of covariance terms on the standard error, it is useful to calculate the standard error using (inappropriately in this example) the assumption of independence. Using the standard errors reported in Table 4 and assuming total independence, the standard error would then be as follows:

$$Gain - Model\ Composite\ Standard\ Error$$
$$= \sqrt{\begin{array}{l}\left(\frac{1}{6}\right)^2\left(SE\ Math_{4_{Year\ 1}}\right)^2 + \left(\frac{1}{6}\right)^2\left(SE\ Math_{5_{Year\ 1}}\right)^2 + \left(\frac{1}{6}\right)^2\left(SE\ Math_{4_{Year\ 2}}\right)^2 \\ + \left(\frac{1}{6}\right)^2\left(SE\ Math_{5_{Year\ 2}}\right)^2 + \left(\frac{1}{6}\right)^2\left(SE\ Math_{4_{Year\ 3}}\right)^2 + \left(\frac{1}{6}\right)^2\left(SE\ Math_{5_{Year\ 3}}\right)^2\end{array}} \qquad (28)$$
$$= \sqrt{\begin{array}{l}\left(\frac{1}{6}\right)^2(0.70)^2 + \left(\frac{1}{6}\right)^2(1.00)^2 + \left(\frac{1}{6}\right)^2(0.50)^2 \\ + \left(\frac{1}{6}\right)^2(1.10)^2 + \left(\frac{1}{6}\right)^2(0.60)^2 + \left(\frac{1}{6}\right)^2(0.70)^2\end{array}} = 0.32$$

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 0.32 due to students being in both Math grade 4 and grade 5 in the school with the specific value depending on the values of the correlations between pairs of value-added gains. Correlations of gains across years can be positive or slightly negative as the same student's score can be used in multiple gains. The magnitude of each correlation depends on the extent to which the same students are in both estimates for any two subject, grade, and year estimates.

For the sake of simplicity, let us assume that the actual standard error was 0.50 for the school in this example.

5.2.5 Calculate Growth Standard Methodology-Based Composite Index Across Subjects

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

$$School\ Three - Year\ Average\ Index = \frac{Gain - Model\ Composite\ Gain}{Gain - Model\ Composite\ SE} = \frac{1.68}{0.50} = 3.37 \qquad (29)$$

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

This example showed a three-year average across grade measure for Math. A single subject and grade or course three-year average measure would use the same methodology except it would be safe to assume independence in the standard error calculation.

5.3 Act 13 Building Level Score and Future Ready PA Index Composite Calculation

This section captures calculations for the Act 13 Building Level Score and Future Ready PA Index measures.

5.3.1 Policy Decisions and Business Rules

The key policy decisions can be summarized as follows:

- The index composites are reported by subject area across grades and include growth measures from the following assessments:
  - Math: PSSA Math 4–8 and Keystone Algebra I
  - ELA: PSSA ELA 4-8 and Keystone Literature
  - Science: PSSA 4 and 8 and Keystone Biology
- The index composites only include subject and grade measures from the current year.
- Within each index composite, there are two components, one based on measures from the growth standard methodology, and one based on measures from the predictive methodology. The former is weighted according to the number of grades included in the across-grades measure, and the latter gives a weight of one to each growth measure.
- The index composite for Act 13 Building Level Score uses the same business rules as described in Section 7, including 11 as the minimum number of students.
- The index composite for schools, LEA/districts, and state calculations also uses the same business rules as described in Section 7.
- The index composite for Future Ready PA Index uses the same business rules as described in Section 7, with the exception that there must be at least 20 students (rather than 11 like the Act 13 Building Level Score) included in the calculation to report the score. The number of students used within each component is calculated and summed for each subject area. If this number is less than 20, then that subject area measure is not reported for that school.

5.3.2 Calculations

Calculations for these measures are nearly identical; however, as noted above, Future Ready PA Index measures are not reported if the total number of students is below 20, and the Act 13 Building Level Score uses a threshold of 11 students. This section provides a sample calculation for the Act 13 Building Level Score based on a sample school.

Table 5: Sample Value-Added Information for a School in Math

Year	Subject	Grade	Growth Measure	Standard Error	Average Growth Index
1	Algebra I	N/A	16.81	3.32	5.06
1	Math	7	-1.96	0.86	-2.28
1	Math	8	3.97	1.00	3.97

5.3.2.1 Calculate Growth Standard Methodology Component

The value-added measures for growth standard methodology subjects, PSSA Math and ELA, are incorporated into Act 13 Building Level Score and the Future Ready PA Index as an across-grades measure for each subject. These measures are all reported in NCEs, so that the growth measures and standard errors can be combined across grades. Similar to the steps described above, there are several key steps to calculating the growth standard methodology component:

1. Calculate the growth standard methodology composite gain across grades
2. Calculate the growth standard methodology composite standard error across grades
3. Calculate the growth standard methodology index across grades

Note that this example reports rounded values for display purposes. The actual calculation uses non-rounded values until the final index.

In this example, there are two applicable grades for this subject, PSSA Math in grades 7 and 8. The composite gain is calculated as follows:

$$\begin{aligned} &Growth\ Standard\ Methodology\ Composite\ Gain \\ &= \frac{1}{2}Math_7 + \frac{1}{2}Math_8 \\ &= \left(\frac{1}{2}\right)(-1.96) + \left(\frac{1}{2}\right)(3.97) = 1.01 \end{aligned} \tag{30}$$

The standard error is calculated assuming independence (because within a subject, there are no students contained in multiple grades within a year). This simplifies the calculation to the following:

$$\begin{gathered} Growth\ Standard\ Methodology\ Comp\ SE = \\ \sqrt{\left(\frac{1}{2}\right)^2 (SE\ Math_7)^2 + \left(\frac{1}{2}\right)^2 (SE\ Math_8)^2} = \sqrt{\left(\frac{1}{2}\right)^2 (0.86)^2 + \left(\frac{1}{2}\right)^2 (1.00)^2} = 0.66 \end{gathered} \tag{31}$$

As a final step, the composite index is calculated by dividing the composite gain by the composite standard error:

$$Growth\ Standard\ Methodology\ \ Math\ Across\ Grades\ Index = \frac{Composite\ Gain}{Composite\ SE} = \frac{1.01}{0.66} = 1.52 \tag{32}$$

Because measures for predictive-methodology subjects (Keystone and PSSA Science) are reported in scale scores, they cannot be combined in this way as scale scores do not have the same interpretation across assessments. They will be included at a later step as illustrated below.

5.3.2.2 Calculate the Combined Values to Generate an Act 13 Building Level Score

First, index components within each subject area are converted to a 0-100 scale in accordance with the table below.

Table 6: Act 13 Building Level Conversion Table

If the index is	The 0-100 Scale Score is
3.00 or greater	100
Less than 3.00 but greater than or equal to 1.00	10 x (Index + 7), truncated to a whole number
Less than 1.00 but greater than or equal to -1.00	5 x (Index + 15), truncated to a whole number
Less than -1.00 but greater than or equal to -3.00	10 x (Index + 8), truncated to a whole number
Less than -3.00	50

To calculate the final Act 13 Building Level Score (on 0-100 scale), each component is first assigned a weight. As described above, for growth standard methodology subjects, weight is based on the number of grades receiving a value-added measure that was included in the across grades measure. For predictive methodology subjects, each measure is given a weight of one. The 0-100 scale score is multiplied by the weight to generate a weighted score. Finally, the weighted scores for each component are then summed within each subject area, divided by the total weight, and then rounded to two decimal places to give the final 0-100 Act 13 Building Level Score.

$$\begin{gathered} Growth\ Standard\ Methodology\ Math\ Building\ Level\ Score = \\ 10 \times (Gain\ Math\ Across\ Grades\ Index + 7) = 85 \\ Predictive\ Methodology\ Algebra\ I\ Building\ Level\ Score = 100 \\ Building\ Level\ Score \\ = \left(\frac{2}{3}\right)(Growth\ Standard\ Math\ Building\ Level\ Score) \\ + \left(\frac{1}{3}\right)(Predictive\ Algrebra\ I\ Building\ Level\ Score) = \frac{170}{3} + \frac{100}{3} = 90 \end{gathered} \tag{33}$$

The final Act 13 Building Level Score for this school in Math is 90.

5.4 ESSA Composite Calculation for Designation and Exit Criteria

5.4.1 Policy Decisions and Business Rules

A composite index is created for the most recent year and the most recent two years for ESSA designation. The policy decisions and business rules are the same as Section 5.3.1, with the exception that ESSA designation measure is across subjects, and the exit measure is subject specific. To receive these measures, each subject area and year included must meet the minimum requirement of 20 students.

5.4.2 Calculations

The key steps for determining a school's growth index for ESSA composites for designation are as follows:

1. For growth measures based on the *growth standard methodology*, calculate composite gain, standard error, and index across grades and subjects for the current year and the immediate prior year (for 2-year measures), and the second prior year (for 3-year measures).
2. For growth measures based on the *predictive methodology*, calculate composite gain, standard error, and index across grades and subjects for the current year, immediate prior year (for 2-year measures), and the second prior year (for 3-year measures).
3. Calculate *composite index* using composite indices from both the growth standard and predictive methodologies.

For the purposes of this sample calculation, assume that Table 7 below provides growth measures for a sample school. This example uses three years of data, but the two-year calculation is done the same way without the earliest year. The example also provides evaluated equations in which the result is rounded assuming all prior calculated values were unrounded to that point as described in Section 4.6.

Table 7: Sample Value-Added Information for a School

Year	Subject	Grade	Value-Added Measure	Standard Error	Index
1	Algebra I	N/A	6.2	3.5	1.77
1	Math	6	0.5	0.3	1.67
1	Math	7	0.4	0.4	1.00
1	Math	8	-0.1	0.3	-0.33
1	ELA	6	0.1	0.2	0.50
1	ELA	7	-0.2	0.3	-0.67
1	ELA	8	-1.2	0.3	-4.00
2	Algebra I	N/A	13	5.5	2.36
2	Math	6	0.3	0.3	1.00
2	Math	7	-0.2	0.4	-0.50

Year	Subject	Grade	Value-Added Measure	Standard Error	Index
2	Math	8	-0.3	0.3	-1.00
2	ELA	6	0.3	0.2	1.50
2	ELA	7	-0.2	0.3	-0.67
2	ELA	8	-0.6	0.3	-2.00
3	Algebra I	N/A	11.2	2.3	4.87
3	Math	6	0.1	0.2	0.50
3	Math	7	-0.5	0.5	-1.00
3	Math	8	0.2	0.5	0.40
3	ELA	6	1.1	0.4	2.75
3	ELA	7	-0.8	0.3	-2.67
3	ELA	8	1.5	0.4	3.75

5.4.3 Calculate Growth Standard Methodology Composite Index Across Subjects

The first step for ESSA designation criteria is similar to what was done for Section 5.2 above or the growth standard methodology calculations in Section 5.3 except that it would be based on the growth measures across all growth standard methodology subjects (rather than just one), and it would also be based on both the most recent two and three years of growth measures. The example below describes the process for creating the three-year designation growth standard methodology composite. As mentioned previously, this example reports rounded values for display purposes. The actual calculation uses non-rounded values until the final index.

The ESSA composite gain is a simple average of the 18 growth standard methodology value-added measures in the two subjects (Math and ELA). More specifically, the gain is calculated using the following formula:

$$Composite\ Gain\ for\ Growth\ Standard\ Methodology = \frac{1}{18}Math_{6_{Year\ 1}} + \frac{1}{18}Math\_7_{\ Year\ 1} + \frac{1}{18}Math_{8_{Year\ 1}} + \frac{1}{18}ELA_{6_{Year\ 1}} + \frac{1}{18}ELA_{7_{Year\ 1}} + \frac{1}{18}ELA_{8_{Year\ 1}} + \frac{1}{18}Math\_6_{\ Year\ 2} + \frac{1}{18}Math_{7_{Year\ 2}} + \frac{1}{18}Math_{8_{Year\ 2}} + \frac{1}{18}ELA_{6_{Year\ 2}} + \frac{1}{18}ELA_{7\ Year\ 2} + \frac{1}{18}ELA_{8_{Year\ 2}} + \frac{1}{18}Math_{6_{Year\ 3}} + \frac{1}{18}Math_{7_{Year\ 3}} + \frac{1}{18}Math_{8_{Year\ 3}} + \frac{1}{18}ELA_{6_{Year\ 3}} + \frac{1}{18}ELA\_7_{\ Year\ 3} + \frac{1}{18}ELA_{8_{Year\ 3}} \qquad (34)$$

$$= \left(\frac{1}{18}\right)(0.5) + \left(\frac{1}{18}\right)(0.4) + \left(\frac{1}{18}\right)(-0.1) + \left(\frac{1}{18}\right)(0.1) + \left(\frac{1}{18}\right)(-0.2) + \left(\frac{1}{18}\right)(-1.2) + \left(\frac{1}{18}\right)(0.3) + \left(\frac{1}{18}\right)(-0.2) + \left(\frac{1}{18}\right)(-0.3) + \left(\frac{1}{18}\right)(0.3) + \left(\frac{1}{18}\right)(-0.2) + \left(\frac{1}{18}\right)(-0.6) + \left(\frac{1}{18}\right)(0.1) + \left(\frac{1}{18}\right)(-0.5) + \left(\frac{1}{18}\right)(0.2) + \left(\frac{1}{18}\right)(1.1) + \left(\frac{1}{18}\right)(-0.8) + \left(\frac{1}{18}\right)(1.5) = 0.02$$

Again, the gains in this composite are not independent since the same students are represented in multiple value-added measures. The steps taken to calculate the standard error are similar to the approach shown in Section 5.2.4. Under the assumption of total independence, the calculation is as follows:

$$Composite\ Standard\ Error\ for\ Growth\ Standard\ Methodology =$$
$$\sqrt{\left(\frac{1}{18}\right)^2 (SE\ Math\ 6_{Year\ 1})^2 + \left(\frac{1}{18}\right)^2 (SE\ Math\ 7_{Year\ 1})^2 + \cdots + \left(\frac{1}{18}\right)^2 (SE\ ELA\ 8_{Year\ 3})^{\ 2}}$$
$$= \sqrt{\left(\left(\frac{1}{18}\right)^2 \left(\begin{matrix} 0.3^2 + 0.4^2 + 0.3^2 + 0.2^2 + 0.3^2 + 0.3^2 + 0.3^2 + 0.4^2 + 0.3^2 + 0.2^2 + 0.3^2 + 0 \\ +0.2^2 + 0.5^2 + 0.5^2 + 0.4^2 + 0.3^2 + 0.4^2 \end{matrix}\right.\right.} \tag{35}$$
$$= 0.08$$

As before, the actual standard error will likely be above the value of 0.08 since there is correlation of measures between subjects and across years that share the same students. In this example, after accounting for correlation, we find that the actual standard error is 0.10. Then, the growth standard methodology index value can be obtained by dividing the average composite gain by the standard error.

$$Composite\ Index\ for\ Growth\ Standard\ Methodology = \frac{Composite\ Gain}{Composite\ SE} = \frac{0.022}{0.10} \tag{36}$$
$$= 0.22$$

5.4.4 Calculate Predictive-Methodology Index Across Subjects

For our sample school, there are three value-added measures from the predictive methodology, and they are based on Algebra I. This composite index weights each individual measure equally.

$$Composite\ Gain\ for\ Predictive\ Methodology = \left(\frac{1}{3}\right)(6.2) + \left(\frac{1}{3}\right)(13.0) + \left(\frac{1}{3}\right)(11.2) \tag{37}$$
$$= 10.13$$

Calculating the predictive-methodology component of the overall index is done the same way as in the growth standard methodology calculations above; however, we assume independence for the standard error because there will likely not be the same student in Algebra I in multiple years. In this example, the standard error would be as follows and not assumed to be higher as in the gain-model example:

$$Composite\ Standard\ Error\ for\ Predictive\ Methodology =$$
$$\sqrt{\left(\frac{1}{3}\right)^2 (SE\ Algebra\ I_{Year\ 1})^2 + \left(\frac{1}{3}\right)^2 (SE\ Algebra\ I_{Year\ 2})^2 + \left(\frac{1}{3}\right)^2 (SE\ Algebra\ I_{Year\ 3})^{\ 2}}$$
$$= \sqrt{\left(\left(\frac{1}{3}\right)^2 (3.5^2 + 5.5^2 + 2.3^2)\right)} = 2.30 \tag{38}$$

The final predictive methodology index is then 10.13 divided by 2.30, or 4.40.

5.4.5 Calculate the Combined Growth Standard and Predictive Methodology Composite Index Across Subjects

The two composite indices from the growth standard and predictive methodologies are then weighted according to the number of value-added measures associated with each model to determine the combined composite index. Our sample school has 21 growth measures, of which 18 are based on the growth standard methodology and three are based on the predictive methodology, so the combined composite index would be calculated as follows using these weightings, the non-rounded growth standard methodology composite index across subjects, and the non-rounded predictive-methodology index across subjects:

$$Unadjusted\ Combined\ Composite\ Index = \left(\frac{18}{21}\right)(0.22) + \left(\frac{3}{21}\right)(4.40) = 0.82 \tag{39}$$

As with the predictive methodology composite, this combined index must be divided by its standard error. In this example, the standard error would be as follows:

$$Final\ Combined\ Composite\ Standard\ Error = \sqrt{\left(\frac{18}{21}\right)^2 (1)^2 + \left(\frac{3}{21}\right)^2 (1)^2} = 0.87 \tag{40}$$

Therefore, the final combined composite index value using non-rounded values is 0.82 divided by 0.87, or 0.92. This is the value that determines the school composite for the ESSA 3-year Designation.

5.5 Available Composites

The following table summarizes the available composites for ESSA Report Card, Future Ready, and designation criteria. As a point of comparison, the table also includes composites from the PVAAS restricted site and Building Level Score.

Table 8: List of Composites by Type

Level	Subjects Included	Number of Years	Student Groups Included
ESSA Report Card/Future Ready			
State	Subject Area measure for Math, ELA, and Science	One	All students; American Indian/Alaskan Native; Asian; Black; Combined Ethnicity; Economically Disadvantaged; English Learners; Foster; Hispanic; Hawaiian/Pacific Islander; Homeless; Military-Connected Family; Students with IEPs; Two or More Races; and White
District	Subject Area measure for Math, ELA, and Science	One	All students; American Indian/Alaskan Native; Asian; Black; Combined Ethnicity; Economically Disadvantaged; English Learners; Foster; Hispanic; Hawaiian/Pacific Islander; Homeless; Military-Connected Family; Students with IEPs; Two or More Races; and White

Level	Subjects Included	Number of Years	Student Groups Included
School	Subject Area measure for Math, ELA, and Science	One	All students; American Indian/Alaskan Native; Asian; Black; Combined Ethnicity; Economically Disadvantaged; English Learners; Foster; Hispanic; Hawaiian/Pacific Islander; Homeless; Military-Connected Family; Students with IEPs; Two or More Races; and White
2-Year Composite for Designation Criteria			
School	Across Math and ELA measure	Two	All students; American Indian/Alaskan Native; Asian; Black; Combined Ethnicity; Economically Disadvantaged; English Learners; Foster; Hispanic; Hawaiian/Pacific Islander; Homeless; Military-Connected Family; Students with IEPs; Two or More Races; and White
3-Year Composite for Designation Criteria (Not Available for 2022-23)			
School	Across Math and ELA measure	Three	All students; American Indian/Alaskan Native; Asian; Black; Combined Ethnicity; Economically Disadvantaged; English Learners; Foster; Hispanic; Hawaiian/Pacific Islander; Homeless; Military-Connected Family; Students with IEPs; Two or More Races; and White
PVAAS Restricted Site			
District	Subject Area measures for Math, ELA, and Science	One	American Indian/Alaskan Native; Asian; Black; Economically Disadvantaged; English Learners; Hispanic; Hawaiian/Pacific Islander; Students with IEPs; Two or More Races; White; Lowest Performing 33% of Students; and Students with GIEPs
PVAAS Restricted Site			
School	Subject Area measures for Math, ELA, and Science	One	American Indian/Alaskan Native; Asian; Black; Economically Disadvantaged; English Learners; Hispanic; Hawaiian/Pacific Islander; Students with IEPs; Two or More Races; White; Lowest Performing 33% of Students; and Students with GIEPs
Building Level Score			
School	Subject Area measures for Math, ELA, and Science	One	All students

6 Input Data Used in the Pennsylvania Growth Model

6.1 Assessment Data Used in Pennsylvania

6.1.1 State Assessments

For the analysis and reporting based on the 2022-23 school year, EVAAS received the following state assessments for use in the PVAAS growth and/or projection models:

- PSSA Mathematics and ELA in grades 3–8
- PSSA Science in grades 4 and 8
- Keystone Algebra I, Biology, and Literature
- ACCESS for ELLs in Composite, Listening, Reading, Speaking, and Writing in grades K-12
- ACCESS for ELLs - Alternate in Composite, Listening, Reading, Speaking, and Writing in grades K-12
- PASA Math and ELA in grades 3–8, 11 (reported only in Student Testing History)
- PASA Science in grades 4, 8, and 11 (reported only in Student Testing History)

PSSA assessments are administered in the spring semester of the school year, and Keystones are administered in the summer, winter, and spring terms (only winter and spring terms in 2020-21). ACCESS for ELLs is administered in the fall and spring terms.

6.1.2 National Assessments

EVAAS also received the following national assessments:

- PSAT 8/9 in Mathematics and Evidence-Based Reading and Writing
- PSAT 10 in Mathematics and Evidence-Based Reading and Writing
- PSAT NMSQT in Mathematics and Evidence-Based Reading and Writing
- SAT in Mathematics and Evidence-Based Reading and Writing
- ACT English, Mathematics, Reading, and Science
- AP Biology, Calculus AB, English Language and Composition, English Literature and Composition, Psychology, Statistics, United States Government and Politics, United States History

The PSAT 10 and AP assessments are administered in the spring semester of the school year. ACT, PSAT 8/9, and SAT are administered in the fall and spring terms. The PSAT NMSQT is administered in the fall.

6.1.3 Locally Administered Assessments

PVAAS received additional local assessments from LEAs that opted to submit them through the Pennsylvania Information Management System (PIMS) Student Local Assessment Subtest Template for additional LEA/district or school reporting. These assessments are often administered each year at the beginning of year (BOY), middle of year (MOY) and end of year (EOY). BOY includes test scores from August 1 through November 14, MOY includes test scores from November 15 through March 14, and EOY includes test scores from March 15 through June 30. The local assessments that were submitted from LEAs include the following:

- Math
  - Acadience (DIBELS Next) Math Composite in grades K-6
  - AimsWeb Math Composite in grades K-8
  - EasyCBM (Total Math) in grades K-8
  - Exact Path Math in grades K-8
  - Fast Bridge Math in grades K-10
  - Fast Bridge Early Math
  - Imagine Learning Math
  - i-Ready (Math Overall) in grades K-8
  - IXL Overall Math
  - Link It Math
  - MAP Math (English or Spanish version) in grades K-12
  - STAR Math in grades 1-12
- English Language Arts
  - Acadience ELA Composite in grades K-6
  - Acadience Oral Reading Fluency in grades 2-6
  - AimsWeb ELA Composite in grades K-8
  - AimsWeb (Oral Reading Fluency) in grades 1-8
  - DIBELS English Language Arts Composite in grades K-6
  - DIBELS English Language Arts Oral Reading Fluency in grades 1-6
  - DRA ELA Comprehension – Independent Level
  - Easy CBM Passage Reading Fluency in grades 2-8
  - Easy CBM Multiple Choice Reading Comprehension in grades 2-8
  - Exact Path English Language Arts in grades K-8
  - Exact Path Reading in grades K-8
  - FastBridge Reading
  - FastBridge early Reading – English
  - Fountas & Pinell ELA Text Level
  - Imagine Learning ELA Literacy
  - Imagine Learning ELA Language
  - i-Ready Reading Overall Score in grades K-8
  - IXL Overall Language Arts
  - Link It English Language Arts Total
  - MAP Language (English or Spanish version)
  - MAP Reading (English or Spanish version) in grades K-12
  - STAR Early Literacy
  - STAR Reading in grades K-12
  - STAR CBM Passage Oral Reading

PDE submitted Classroom Diagnostic Tests (CDTs) on behalf of LEAs, and this includes the following content areas for BOY and MOY timepoints:

- Algebra I
- Algebra II
- Biology
- Chemistry

- Geometry
- Math Grades 3–12
- Reading/Literature 3-12
- Science Grades 3–12
- Writing/English Composite 3-12

For CDTs, BOY includes test scores from August 1 through November 14, MOY includes test scores from November 15 through March 14, and EOY incudes test scores from March 15 through June 30.

State assessment files from PDE provided the following data for each student score:

- Student last name
- Student first name
- Middle initial (if available)
- Student date of birth
- PA secure ID
- Scale score
- Performance level
- Test taken
- Tested grade
- Tested semester
- District AUN
- School code

Some of this information, such as performance levels, is not relevant to PSAT or SAT tests. Similar data was collected for local assessments and CDTs. No student scores are utilized from any non-public schools. These include prior year scores for students who are now in public schools.

6.2 Student Information

Student information is used in creating the web application to assist educators in analyzing the data to inform practice and assist all students with academic growth. SAS receives this information in the form of various socioeconomic, demographic, and programmatic identifiers provided by PDE. SAS received the following student information and identifiers from PDE:

- English Learner
- English Learner (EL) First Year
- English Learner (Federal)
- Economically Disadvantaged
- Enrolled Full Year
- Foreign Exchange
- Foster
- Gender
- Gifted Education (GIEP) – added with enrollment data
- Homeless
- Homeschool

- Individualized Education Plan (IEP)
- Migrant
- Military-Connected Family
- Race
  - American Indian/Alaskan Native
  - Asian
  - Black
  - Economically Disadvantaged
  - English Learner
  - Hawaiian/Pacific Islander
  - Hispanic
  - Two or More Races
  - White
- Service Plan 504 – added with enrollment data
- Title 1
- Title 3

6.3 Teacher Information

In order to provide accurate and verified student-teacher linkages in the teacher growth models, Pennsylvania educators, school leaders, and LEA/district leaders 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.

EVAAS receives data from the PIMS Staff Student Subtest Template provided by PDE that contains a record for each teacher/student instructional relationship for each assessment. The student-teacher linkage files include the following information:

- District AUN
- District name
- School code
- School name
- Teacher level identification
  - Teacher name
  - PPID
- Student linking information, including PA secure ID
- Subjects
- Semester
- Percentage of concurrent enrollment
- Percentage of shared instruction

As teachers across the Commonwealth participate in roster verification, the last two pieces of information are modified as needed. The first is the percentage of concurrent enrollment, which is the percentage that a teacher and student are concurrently enrolled with one another from day one of the subject and grade or course through the last instructional day before the LEA/district's testing window for that subject and grade or content area. The second is the percentage of shared instruction, which captures information about team teaching or shared instruction between two or more teachers. These two percentages are determined by LEAs/districts. Both pieces of information are multiplied together to obtain an overall percentage of instructional responsibility.

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 growth can be measured for both students with histories of lower achievement as well as students with histories of higher achievement. A floor or ceiling in the scales could disadvantage educators serving either students with lower achievement or students with higher achievement.
2. **The test is highly related to the academic standards** so that it is possible to measure growth 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 Pennsylvania's current state standardized assessments meet them. Only local assessments and CDTs that met these requirements received LEA/district or school reporting. 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 Pennsylvania 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 have histories of high or low achievement.

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. This is how state tests are designed and is monitored by the PDE and their psychometricians.

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. The type of reliability is important for most any use of standardized assessments. This criterion typically is met when there is a sufficient number of items per subject/grade/year, and this will be monitored each subsequent year that the test is given.

7.2 Pre-Analytic Processing

7.2.1 Missing Grade

In Pennsylvania, the grade used in the analyses and reporting is the tested grade, not the enrolled grade. If a grade is missing on an any PSSA tests, 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.

Of the 1,672,817 records from the 2022-23 PSSA Math, ELA, and Science assessments, no records were excluded due to this business rule.

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 score with the most demographic information will be retained for the analysis and reporting, and the other score will be removed from the analysis and reporting.

Of the 2,222,631 records from the 2022-23 PSSA Math, ELA, and Science and Keystone Algebra I, Biology, and Literature assessments, six records (0.0003%) were excluded due to this business rule.

7.2.3 Students with Missing LEAs/Districts or Schools for Some Scores but Not Others

If a student has a score with a missing LEA/district or school for a particular subject and grade in a given testing period, then the duplicate score that has a LEA/district and/or school will be included over the score that has the missing data.

Of the 2,222,631 records from the 2022-23 PSSA Math, ELA, and Science and Keystone Algebra I, Biology, and Literature assessments, no records were excluded due to this business rule.

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

7.2.4.1 State Assessments

For state assessments, 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 all duplicate records will be excluded from the analysis.

Of the 2,222,631 records from the 2022-23 PSSA Math, ELA, and Science and Keystone Algebra I, Biology, and Literature assessments, 86 records (0.004%) were excluded due to this business rule.

7.2.4.2 Local Assessments and CDTs

For local assessments and CDTs, if a student has multiple scores in the same period, then the following is used to determine their BOY/MOY/EOY score.

- The BOY test score is defined as the record with the first test date for a student/test/subject/grade.
- The EOY test score is defined as the record with the last test date for a student/test/subject/grade.

- The MOY test score is defined as the record with the earliest test date from December/January. If no records are available then the MOY test score is defined as the earliest test date from November, February, March.

If a student then has multiple records on the same test date in BOY, MOY, or EOY then all such records are excluded.

Of the 2,271,946 records from the 2022-23 local assessments and CDTs, 221,165 records (9.7%) were excluded due to this business rule.

7.2.4.3 College Readiness Assessments

For college readiness assessments, the score with the earliest test date within the school year is used for the analysis. If a student has multiple records on the same test date, then all such records are excluded.

Of the 11,590,123 records from 2022-23 local assessments and CDTs, 221,165 records (1.91%) were excluded due to this business rule.

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. This rule applies to PSSA only.

Of the 1,672,817 records from 2022-23 PSSA Math, ELA, and Science assessments, four records (0.0002%) were excluded due to this business rule.

7.2.6 Students with Records That Have Unexpected Grade Level Changes

If a student skips more than one grade level (e.g., moves from sixth in year 1 to ninth in year 2) or is moved back by one grade or more (i.e., moves from fourth in year 1 to third in year 2) 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.

Of the 1,672,817 records from 2022-23 PSSA Math, ELA, and Science assessments, no records were excluded due to this business rule.

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. In Pennsylvania, it can happen that a student is accelerated in a subject and tests at two different schools.

7.2.8 Students Flagged as Homeschool

If a student is identified as a homeschool student, then the score will be excluded from the analysis and reporting.

7.2.9 Students Flagged as Not Meeting Enrollment Criteria

If a student is flagged as No for full year enrollment, then the score will be excluded from the district and school analysis and reporting.

7.2.10 Students Flagged as First-Year English Learners in the Current Year

If a student is flagged as a first-year EL in the current year, then the score will be excluded from the analysis and reporting. Note that future scores are eligible to be included.

7.2.11 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 on state assessments, all Math scores from state assessments are examined simultaneously during outlier identification for the state assessments, 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 (e.g., Math subjects with PSSA and Keystones tests might be used to identify outliers with SAT).

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 identified, then that outlier will not be used in the analysis, but it will be displayed on the student testing history on the PVAAS 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 -3.5 when determining the difference between the score in question and the weighted combination of reference scores (otherwise known as the comparison score) for PSSA Math and ELA. In other words, the score in question must be at least 3.5 standard deviations below the comparison score. For other assessments (PSSA Science, Keystones, local assessments, CDTs), 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 4.0 when determining the difference between the score in question and the reference group of scores for PSSA Math and ELA. In other words, the score in question must be at least 4.5 standard deviations above the comparison score. For other assessments (PSSA Science, Keystones, local assessments, CDTs), 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.

Of the 2,222,631 records from the 2022-23 PSSA Math, ELA, and Science and Keystone Algebra I, Biology, and Literature assessments, 1,290 records (0.06%) were excluded due to this business rule.

7.2.12 Linking Records over Time

Each year, PVAAS 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 PPID and teacher's name.

7.3 Growth Models

7.3.1 Students Included in the Analysis

7.3.1.1 LEA/District and School Model

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

For the growth standard methodology, all students are included in these analyses if they have assessment scores that can be used. The growth standard methodology for state assessments uses all available PSSA Math and ELA 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 growth standard methodology 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 growth standard methodology, 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.

For the predictive and projection methodologies, a student must have at least three valid predictor scores (or two scores for selected assessments) that can be used in the analysis, all of which 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 predicted 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 must only have some subset of three that were used in the analysis. Unlike the growth standard methodology, students with non-traditional grade patterns are included in the predictive methodology as one student. Since the predictive methodology 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 due to the fact that 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 methodology, 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 district and school reporting, Keystone scores from the summer, winter, and spring test administrations are considered for use in annual value-added analyses. The summer administration is considered the first administration of the school year, followed by the winter and spring administrations. Only scores from a full administration of the exam (module 1 and 2) are considered for use in the analyses.

The rules for determining which Keystone scores to include in the PVAAS analyses are summarized as follows:

- Starting with the 2020-21 reporting and continuing in the years following as applicable, if a student tested for the first time and was enrolled in a Keystone course in the 2019-20 school year (summer 2019, winter 2019, and spring 2020), then that student score will not be included in the reporting.
- If a student has previously tested in the Proficient or Advanced range in a Keystone subject, no subsequent scores in that same subject will be included in the analyses. In accordance with PDE

policy, students who have already scored at least Proficient should not retest. If applicable, the business rules regarding which retested Keystones scores to include will be applied to these student scores.

- The analysis excludes any student assessment scores in the current year for students who are identified as Special Education in the current year and their district number where they tested does not equal their district number of residence.
- The analysis excludes any student assessment scores in the current year for students who received a higher score on a specific Keystone assessment in a prior year.

All other rules for excluding students' scores from the analyses still apply.

7.3.1.2 Teacher Model

For teacher reporting, Keystone scores from the winter and spring test administrations are considered for use in annual value-added analyses. The summer administration is considered the first administration of the school year, followed by the winter and spring administrations. Only scores from a full administration of the exam (module 1 and 2) are considered for use in the analyses.

The rules for determining which Keystone scores to include in the PVAAS analyses are summarized as follows:

- Starting with the 2020-21 reporting and continuing in the years following as applicable, if a student tested for the first time and was enrolled in a Keystone course in the 2019-20 school year (summer 2019, winter 2019, and spring 2020), then that student score will not be included in the reporting.
- If a student has previously tested in the Proficient or Advanced range in a Keystone subject, no subsequent scores in that same subject will be included in the analyses. In accordance with PDE policy, students who have already scored at least Proficient should not retest. If applicable, the business rules regarding which retested Keystones scores to include will be applied to these student scores.

Additionally, the following rules are applied:

- Student scores are excluded if the students are not claimed by the teacher for at least 10% of their instructional responsibility. Because of this, the full year enrollment designation is not used to exclude students from the analysis.
- Student scores are still excluded if flagged as EL-first year or if the scores are considered outliers.

7.3.1.3 Student Groups Model

Specific to the student groups model, students are identified as members of specific groups by using files with results for the PSSA and Keystone assessments provided to SAS each year. The CDTs, ACCESS for ELLs, and locally administered assessments leverage the same source as the PSSA and Keystones for student identifiers when possible.

PVAAS identifies the student group of the lowest performing 33% of students by using their average or predicted scores. For assessments that use the growth standard methodology, the average is based on the student's current and prior year score in the same subject. For assessments that use the predictive methodology, the predicted score is the same as described in the value-added model section. Students are included in the analysis if the average or predicted score is in the bottom 33% across the school. This bottom 33% is defined in the current (most recent analysis) year for each course using average or predicted scores.

7.3.2 Minimum Number of Students to Receive a Report

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

7.3.2.1 LEA/District and School Model

For the growth standard methodology, the minimum student count to report an estimated average NCE *score* (i.e., either entering or exiting achievement) is 11 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 11 students who are associated with the school or LEA/district in the subject, grade, and year.
- Of those students who are associated with the school or LEA/district in the current year and grade, there must be at least 11 students in each subject, grade, and year in order for that subject, grade, and year to be used in the gain calculation.
- There is at least one student at the school or LEA/district who has a "simple gain," which is based on a valid test score in the current year and grade as well as the prior year and grade in the same subject. Note that, for local assessments, at least 33% of all students must have a simple gain.
- For any LEA/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 PSSA Math grade 5 for this year, there must be the following requirements:

- There must be at least 11 fifth-grade students with a PSSA Math grade 5 score at school A for this year.
- Of the fifth-grade students at school A this year *in all subjects, not just Math*, there must be at least 11 students with a PSSA Math grade 4 score from last year.
- At least one of the fifth-grade students at school A this year must have a PSSA Math grade 5 score from this year *and* a PSSA Math grade 4 score from last year.

For the predictive methodology, the minimum student count to receive a growth measure is 11 students in a specific subject, grade, and year. These students must have the required three prior test scores needed to receive a predicted 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 11 students with a valid test score in the same subject, grade, and year. This requirement does not consider the percentage of instructional responsibility that the teacher has for each student in a specific subject and grade.

The teacher predictive *model* includes teachers who are linked to at least 11 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 responsibility that the teacher has for 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 for the state assessments. The teacher's number of FTE students is based on the number of students linked to that teacher and the percentage of instructional responsibility the teacher has for each student. For example, if a teacher taught 11 students with 50% instructional responsibility for each student, then the teacher's FTE number of students would be 5.5, and the teacher would not receive a teacher growth report. If another teacher taught 14 students with 50% instructional responsibility for each student, then that teacher would have seven FTE students and would receive a teacher growth report. The instructional responsibility attribution is obtained from the student-teacher linkage data described in Section 6.3 and Section 7.4.

The teacher gain model has an additional requirement to ensure there is sufficient student data linked to a teacher. The teacher must be linked to at least five students with prior test score data in the same subject, and the test data can come from any prior grade as long as they are part of the student's regular cohort. 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 for state assessments. Students are linked to a teacher based on the subject area taught and the assessment taken. Students that 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 Model

The minimum number of students required for the calculation of a value-added measure for a student group is the same as the general PVAAS value-added reporting. There must be at least 11 students with sufficient testing history in a specific student group who have taken a grade and subject-specific assessment in a specific year. For any across-grades measure displayed on the web reporting, there must be at least 11 students with sufficient testing history in each subject and grade or Keystone assessment used in that across-grades or overall measure in that specific year.

7.4 Student-Teacher Linkages

The instructional responsibility attribution is obtained from the linkage roster verification process that is in use in PVAAS. This information is outlined in Section 6.3. Students are linked to a teacher based on the subject, grade, and/or course taught and the state assessment taken. In some cases, the course being taught might not directly align to all state assessments taken by the student, and, in those cases, linkage by EVAAS is not mandatory in accordance with PDE policy.

For example, all eighth-grade students take the PSSA Mathematics grade 8 test. However, some eighth-grade students (as well as students in younger grades) are also enrolled in a Keystone Algebra I course in a PSSA-tested grade rather than the general grade 8 Math course. Their teachers will not be automatically linked in the PVAAS roster verification system to these eighth-grade students enrolled in Algebra I by EVAAS unless the LEA submits these links to the PSSA Math assessment into the PIMS Staff-Student-Subtest template. As a result, these teachers might not receive a PSSA Math grade 8 report.

LEAs are responsible for making this determination. LEAs make the choice as to whether to have their teachers linked to such students if the students should be included in the teacher's PSSA Math grade 8 value-added report. If the LEA determines that the Algebra I teacher also has responsibility for eighth-grade students on the grade 8 assessment (or grade 7, etc.), then that teacher would receive a grade 8 mathematics report based on the students who took the PSSA Math Assessment *and* a separate Algebra I report through both state assessments.

The process for creating an accurate link between students and teachers (roster verification) allows teachers and principals to review the attribution used in the PVAAS reports. For more information about teacher roster verification, email pdepvaas@iu13.org.

Student-teacher linkages are connected to assessment data based on the subject and identification information described in Section 6.3. More specifically, the following steps are used:

- PVAAS calculates the percentage of instructional responsibility for each student-teacher linkage, and this metric is calculated as the percentage of concurrent enrollment multiplied by percentage of shared instruction, divided by 100.
- PVAAS excludes students claimed at less than 10% for instructional responsibility in the assessed content area or less from the teacher in the current year.
- 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 *or* course. 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). After adjusting, PVAAS excludes again the students claimed at less than 10% for instructional responsibility in the assessed content area or less from the teacher in the current year.
- In the teacher models, students are weighted according to the percentage of instructional responsibility for a specific teacher, test, subject and grade *or* course.