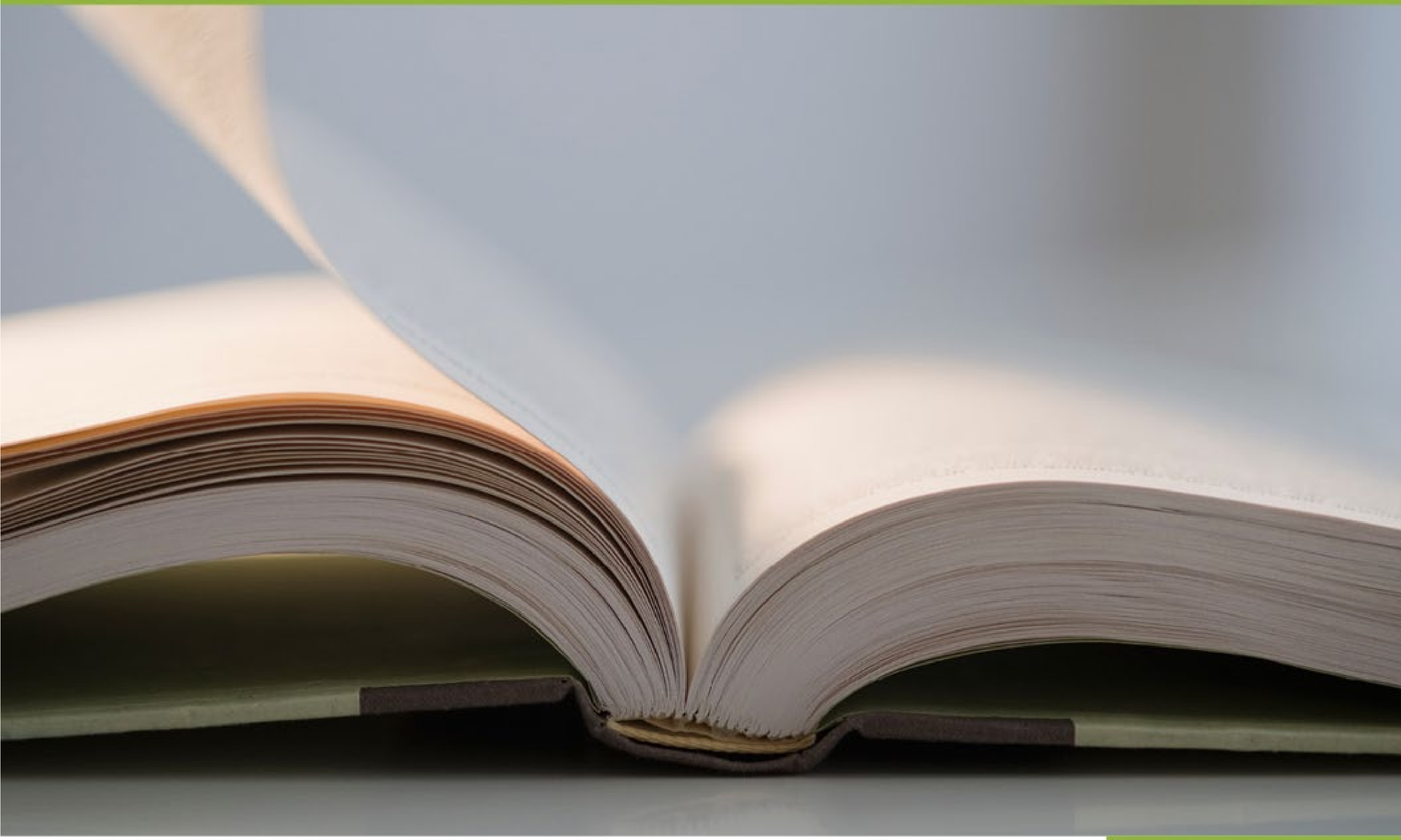
SAS® EVAAS

Statistical Models and Business Rules

Prepared for the Virginia Department of Education



SAS THE POWER TO KNOW.

Contents

1 Introduction to Virginia's Growth Reporting

VVAAS growth models use established statistical analysis to measure students' academic growth. Conceptually and as a simple explanation, 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 intended use of the data, students included in the analysis, and level of reporting (division or school). For each of these models, there are business rules to ensure the growth measures reflect the policies and practices selected by the Commonwealth of Virginia.

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

Third, the growth measures are relative to students' expected growth, which is in turn determined by the growth that is observed within the actual population of Virginia 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 Virginia's growth modeling based on the statistical models, business rules, policies, and practices selected by the Commonwealth of Virginia and currently implemented by EVAAS. This document includes details and decisions in the following areas:

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

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

2 Statistical Models

2.1 Overview of Statistical Models

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

Virginia'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, which is inherent in all academic assessments, by using multiple prior test scores as predictors for each student.

- **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 students 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.

These advantages provide robust and reliable growth measures to divisions and schools. 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.

VVAAS includes output from the following models within the web application:

- **Predictive model** measures growth by comparing students' actual performance on an assessment to their predicted or expected performance.
- **Projection model** provides a probability of obtaining a particular score or higher on a given assessment for individual students. This is similar to the predictive model except that it is intended as an instructional tool for educators serving students who have not yet taken an assessment.

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

2.2 Predictive Model

2.2.1 Overview

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

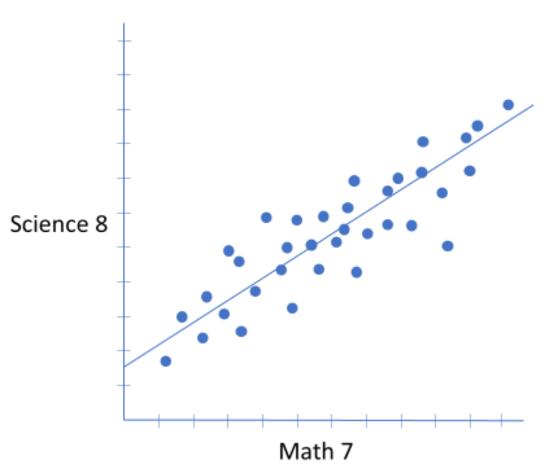
There are two separate analyses for VVAAS reporting based on the predictive model: one for divisions, and another for schools. The division and school models are essentially the same.

Regression models are used in virtually every field of study, and their intent is to identify relationships between two or more variables. When it comes to measuring growth, regression models identify the relationship between prior test performance and actual test performance for a given course. In more technical terms, the predictive model is known as the univariate response model (URM), a linear mixed model and, more specifically, an analysis of covariance (ANCOVA) model.

2.2.2 Conceptual Explanation

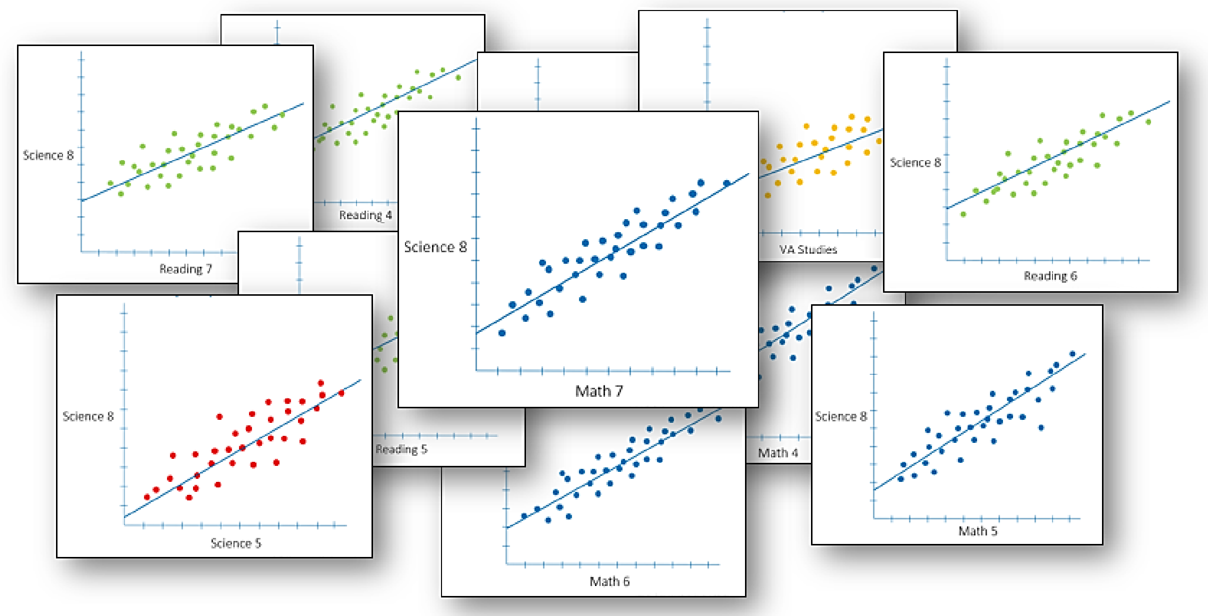
As mentioned above, the predictive model is ideal when previous test performance is used to predict another test performance. Consider all students who tested in SOL Science in grade 8 in a given year. There isn't a Science test in the immediate prior grade. However, these students might have a number of prior test scores in SOL Math and Reading in grades 3–7 as well as SOL Science in grade 5. These prior test scores have a relationship with SOL Science, meaning that how students performed on these tests can predict how the students perform on SOL 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 SOL Math 7 plotted with their score on SOL Science 8. The best-fit line indicates how students with a certain prior score on SOL Math 7 tend to score, on average, on SOL Science 8. This illustration is based on one prior test; the predictive model uses many prior test scores from different subjects and grades.

Figure 1: 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 SOL 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 SOL Science in grade 8 than prior Reading scores, but how a student performs on the Reading test typically provides an idea of how we might expect a student to perform on average on SOL Science. This is shown in Figure 4 below, where there are a number of different tests that have a predictive relationship with SOL Science in grade 8. All of these relationships are considered together in the predictive model, with some tests weighted more heavily than others.

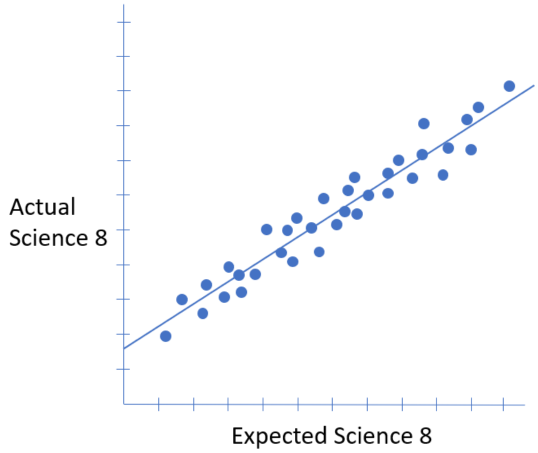
Figure 2: Relationships Observed in the Statewide Data Inform the Predictive Model



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

Each student receives an expected score based on their own prior testing history. In practical terms, the expected score represents the student's entering achievement because it is based on all prior testing information to date. Figure 5 below shows the relationship between expected and actual scores for a group of students.

Figure 3: Relationship Expected Score and Actual Score for Selected Subject and Grade



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

2.2.3 Technical Description of the Division and School Models

The predictive model has similar approaches for divisions and schools. 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 division or school 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.

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

The μ terms are means for the response and the predictor variables. α_j is the school effect for the j^{th} school. The β terms are regression coefficients. Predictions to the response variable are made by using this equation with estimates for the unknown parameters (μ terms and β terms). The parameter estimates (denoted with "hats," e.g., $\hat{\mu}$, $\hat{\beta}$) are obtained using all students that have an observed value for the specific response and have two 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 model. First, not all students will have the same set of predictor variables due to missing test scores. Second, because the predictive model is an ANCOVA model, the estimated parameters are pooled within group (division or school). 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,

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

The *C* 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 (division or school) 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 two predictor variables are included in the estimation. Given such a matrix, the vector of estimated regression coefficients for the prediction 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" division or school is zero), then the appropriate means are the means of the group means. The group-level means are obtained from the EM algorithm mentioned above, which accounts for missing data. The overall means ($\hat{\mu}$ terms) are then obtained as the simple average of the group-level means.

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

The $\hat{y}_i$ term in equation (15) is nothing more than a composite of all the student's past scores. It is a one-number summary of the student's level of achievement prior to the current year, and this term is called the expected score or entering achievement in the web reporting. The different prior test scores making up this composite are given different weights (by the regression coefficients, the $\hat{\beta}$ terms) in order to maximize its correlation with the response variable. Thus, a different composite would be used when the response variable is Mathematics than when it is Reading, for example. Note that the $\hat{\alpha}_j$ term is not included in the equation. Again, this is because $\hat{y}_i$ represents prior achievement before the effect of the current division or school.

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

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

In the predictive model, the effects (α_j) are considered random effects. Consequently, the $\hat{\alpha}_j$ terms are obtained by shrinkage estimation (empirical Bayes).[1] The regression coefficients for the ANCOVA model are given by the γ terms.

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

2.2.3.1 Adjustments based on Student Characteristics

The predictive model makes two types of adjustments:

1. Enrolled grade of a student. The model takes into account the students' enrolled grade for a given assessment, as there could be different enrolled grades for a given assessment. This adjustment also takes into account whether students test in "accelerated" subjects or grades and when the acceleration happens, such as a sixth grade student who takes the SOL Math 7 assessment.
2. English Learners' status of a student.

Accelerated students are primarily observed in SOL Math for grades 5-8. In these assessments, the adjustment is based on the relationship among groups of students, which are based on the enrolled grade. If the enrolled grade group has fewer than 5,000 students, then it will be combined with the next closest enrolled grade group. The enrolled grade group will combine with a higher grade if it is a low grade or a lower grade if it is a high grade (if one exists).

For SOL Math 7 growth measures, there are three groups of students in the model:

- Students enrolled in grade 7 or higher
- Students enrolled in grade 6 who were first accelerated in grade 6
- Students enrolled in grade 6 or less who were first accelerated in grade 5 or earlier

For SOL Math 8 growth measures, there are four groups of students in the model:

- Students enrolled in grade 8 or higher
- Students enrolled in grade 7 who were first accelerated in grade 7
- Students enrolled in grade 7 or earlier who were first accelerated in grade 6 or earlier
- Students enrolled in grade 6 or less

Again, the model considers the relationships among these student groups for a specific assessment, such as SOL Math 7 or SOL Math 8, and adjusts so that there is not an advantage or disadvantage to the type of student group served by the school.

---

[1] 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).

The model also makes adjustments for students identified as English Learners (EL) for the following assessments and test-takers:

- SOL Reading in grades 4-8
- SOL Math in grades 4 and 5
- SOL Math Grade 6 for students enrolled in grade 6 or higher
- SOL Math Grade 7 for students enrolled in grade 7 or higher
- SOL Math Grade 8 for students enrolled in grade 8 or higher
- SOL Science Grade 5
- SOL Science Grade 8 for students enrolled in grade 8 or higher
- SOL VA Studies Grade 4

The accelerated and EL adjustments are made together to determine the response variable. The following is an example of how these adjustments are made in the model for SOL Math Grade 8. Because of the unique students' characteristics, there are five groups of students in the model:

- Math 8 students who are not accelerated (enrolled grade 8 or higher) and are identified as EL
- Math 8 students who are not accelerated (enrolled grade 8 or higher) and are not identified as EL
- Math 8 students who are enrolled in grade 7 and were accelerated first in grade 7
- Math 8 students who are enrolled in grade 7 or earlier and were accelerated first in grade 6 or earlier
- Math 8 students who are enrolled in grade 6 or less

Collectively, the following adjustments are made based on student characteristics. Note that there can be more than one adjustment per assessment, depending on the enrolled grade, accelerated status or English Learner status of the students.

Table 1: Adjustments based on Student Characteristics by Assessment

Assessment	Adjustment based on Enrolled Grade	Adjustment based on Grade when First Accelerated	Adjustment based on English Learners = Yes
Mathematics 4			Yes
Mathematics 5			Yes
Mathematics 6	5		
Mathematics 6	6		Yes
Mathematics 7	6	5	
Mathematics 7	6	6	
Mathematics 7	7		Yes
Mathematics 8	6		
Mathematics 8	7	6	
Mathematics 8	7	7	
Mathematics 8	8		Yes
Reading 4			Yes
Reading 5			Yes
Reading 6			Yes
Reading 7			Yes
Reading 8			Yes
Science 5			Yes
Science 8	7		
Science 8	8		Yes
Writing 8			
VA Studies	4		Yes
VA Studies	5		
Civics and Economics	7		
Civics and Economics	8		
EOC Algebra I	7, 8, 9, 10		
EOC Algebra II			
EOC Biology	9, 10, 11		
EOC Earth Science	8, 9		
EOC World Geography			
EOC Geometry	8, 9, 10		
EOC World History I			
EOC World History II			
EOC Reading	11, 12		
EOC Writing	10, 11, 12		

Equations 15, 17, and 18 above are all unique to these groups of students. In other words, expected scores are created differently for each of these groups. Those expected scores assume the average experience of the students within that group. After those expected scores are calculated, equation 18 is used to create school effects with expected scores from all groups included in this modeling step.

2.2.3.2 Accommodations to the Predictive Model for Missing 2019-20 Data due to the Pandemic

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 SOL assessments based on the 2019-20 school year.

The predictive model is used to measure growth for assessments, regardless of whether they are given in consecutive or non-consecutive grades. With the latter type of assessment, it is always 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 model did not require any technical adaptations to account for the missing year of data.

2.3 Projection Model

2.3.1 Overview

The longitudinal data sets used to calculate growth measures for groups of students can also provide individual student projections to future assessments. A projection is reported as a probability of obtaining a specific score or above on an assessment, such as a 70% probability of scoring Pass/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.

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

2.3.2 Technical Description

The statistical model that is used as the basis for the projections is, in traditional terminology, an analysis of covariance (ANCOVA) model. This model is the same statistical model used in the predictive model applied at the school level described in Section 2.2.3. 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:

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

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

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

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

Two difficulties must be addressed to implement the projections. First, not all students will have the same set of predictor variables due to missing test scores. Second, because this is an ANCOVA model with a school effect i, the regression coefficients must be "pooled-within-school" regression coefficients. The strategy for dealing with these difficulties is the same as described in Section 2.2.3 using equations (15), (16), and (17) and will not be repeated here.

Typically, the parameter estimates are based on the cohort of students who most recently took the assessment. For the 2021-22 reporting, the parameter estimates are based on the 2019 cohort of students since that represents a typical pre-pandemic relationship among students' prior test scores. 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 two available predictor scores. 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 Pass/Proficient on a future grade-level or EOC test. The probability is calculated as the area above the benchmark cutoff score using a normal distribution with its mean equal to the projected score and its standard deviation equal to the standard error of the projected score as described below. Φ represents the standard normal cumulative distribution function.

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

2.4 Outputs from the Models

The outputs of the value-added model are available to Virginia educators with user credentials in the VVAAS web application available at https://VVAAS.sas.com/.

2.4.1 Predictive Model

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

SOL:

- Reading in grades 4-8
- Math in grades 4-8
- Science in grades 5 and 8
- Writing in grade 8
- Civics and Economics
- VA Studies

SOL EOC:

- EOC Algebra I
- EOC Algebra II
- EOC Biology
- EOC Earth Science
- EOC Geometry
- EOC Reading
- EOC World Geography
- EOC World History I
- EOC World History II
- EOC Writing

Collectively, the different models provide metrics for a variety of purposes within the Commonwealth of Virginia. These metrics include division growth measures and school growth measures.

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

2.4.2 Projection Model

Projections are provided to future state assessments. More specifically, most SOL grade-level projections are provided only to a student's next tested grade-level SOL assessments based on that student's most recent tested grade, such as projections to grade 5 for students who most recently tested in grade 4. Science, Writing, Civics and Economics, and VA Studies are provided one or two grades in advance, and EOC projections typically start with students who last tested in grade 7. Projections are made to the performance levels Fail/Basic, Pass/Proficient, and Pass/Advanced depending on the assessment, and the individual cut scores depend on each subject and grade. To summarize, the following SOL projections are available for students who meet the reporting criteria:

- SOL Math and Reading in grades 4–8
- SOL Science in grades 5 and 8
- SOL Writing in grade 8
- SOL Civics and Economics
- SOL VA Studies
- SOL EOC Algebra I, Algebra II, Geometry, Earth Science, Biology, World Geography, World History I, World History II, Reading, Writing

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.

3.2 Technical Description

The predictive model defines expected growth based on the empirical student testing data; in other words, the model does not assume a particular amount of growth or assign expected growth in advance of the assessment being taken by students. The model defines 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 predictive model, expected growth means that students with a division or school made the same amount of growth as students with the average division or school in the Commonwealth for that same year, subject, and grade**.

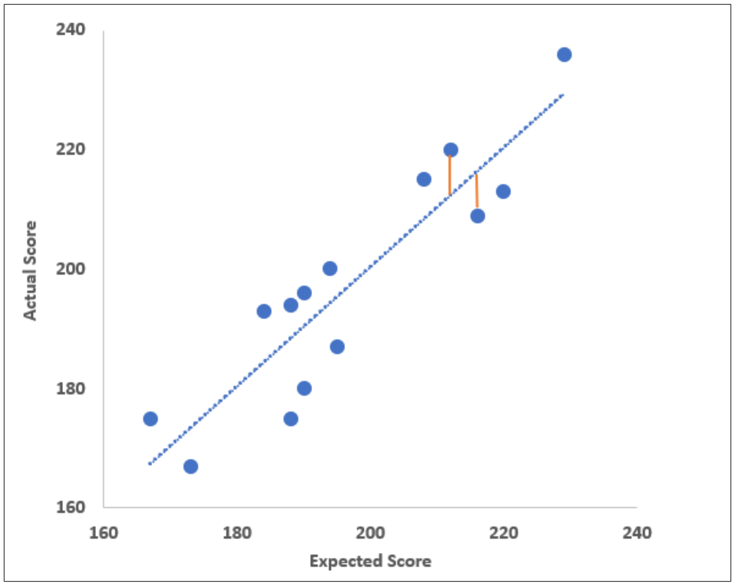
The growth measures tend to be centered on expected growth every year with approximately half of the division/school estimates above zero and approximately half of the division/school estimates below zero.

A change in assessments or scales from one year to the next does not present challenges to calculating expected growth. The predictive model already uses prior test scores from different scales to calculate the expected score. When assessments change over time, expected growth is still based on the relative change in achievement from one point in time to another.

3.3 Illustrated Example

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

Figure 4: Intra-Year Approach Example for the Predictive Model



4 Classifying Growth into Categories

4.1 Overview

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

4.2 Use Standard Errors Derived from the Models

As described in the modeling approaches section, the growth model provides an estimate of growth for a division or school 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. Standard error is a common statistical metric reported in many analyses and research studies because it yields important information for interpreting an estimate, which is, in this case, the growth measure relative to expected growth. Because measurement error is inherent in any growth or value-added model, *the standard error is a critical part of the reporting*. **Taken together, the growth measure and standard error provide educators and policymakers with critical information about the certainty that students in a division or school 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 school as ineffective when it is truly effective).

The standard error also takes into account that even among schools with the same number of students, schools 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 schools, depending on the available data that is associated with their students, and it is another important protection for divisions and schools to incorporate standard errors to the value-added reporting.

4.3 Define Growth Indicators in Terms of Standard Errors

Common statistical usage of standard errors indicates the precision of an estimate and whether that estimate is statistically significantly different from an expected value. The growth reports use the standard error of each growth measure to determine the statistical evidence that the growth measure is different from expected growth. For VVAAS growth reporting, there are thresholds at one and two standard errors above expected growth as well as one or two standard errors below expected growth. These thresholds can also be expressed in terms of a growth index. These definitions then map to growth indicators in the reports themselves, such that there is statistical meaning in these categories. The categories and definitions are illustrated in the following section.

4.4 Illustrated Examples of Categories

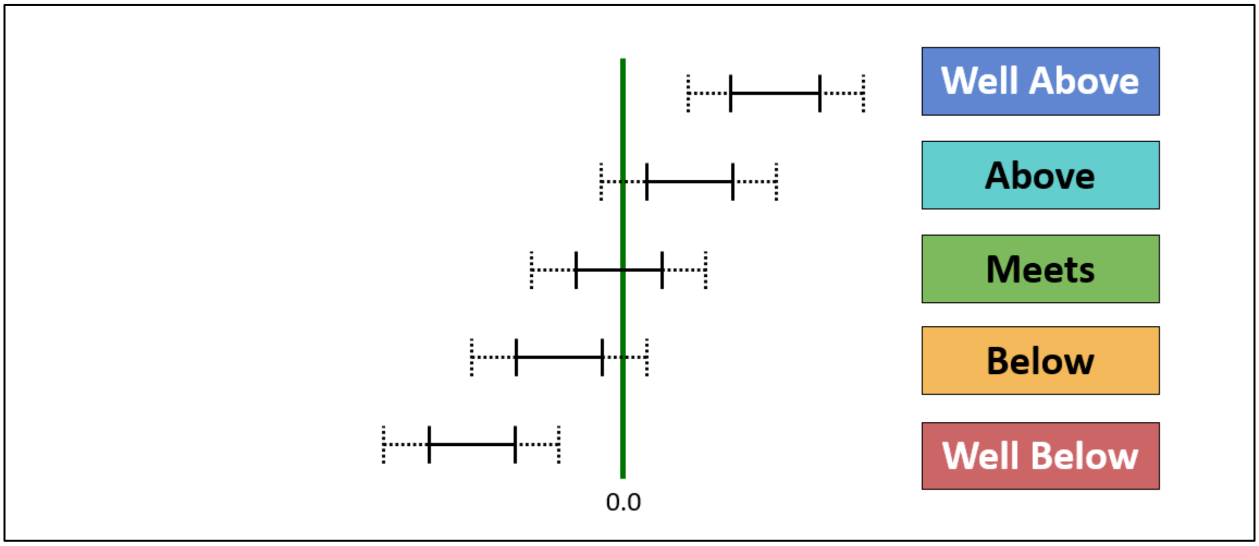
There are two ways to visualize how the growth measure and standard error relate to expected growth and how these can be used to create categories.

The first way is to frame the growth measure relative to its standard error and expected growth at the same time. For division and school 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 level of certainty is significant evidence that students made more growth than expected.
- **Above** indicates that the growth measure is at least one but less than two standard errors above expected growth (0). This is moderate evidence that students made more growth than expected.
- **Meets** indicates that the growth measure is less than one standard error above expected growth (0) but no more than two standard errors below expected growth (0). This is evidence that students made growth as expected.
- **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 that students made less growth than expected.
- **Well Below** is an indication that the growth measure is less than or equal to two standard errors below expected growth (0). This level of certainty is significant evidence that students made less growth than expected.

Figure 8 below shows visual examples of each category. The green line represents the expected growth. The solid black line represents the range of values included in the growth measure plus and minus *one* standard error. The dotted black line extends the range of values to the growth measure plus and minus *two* standard errors. If the dotted black line is completely above expected growth, then there is significant evidence that students made more than expected growth, which represents the 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 5: Visualization of Growth Categories with Expected Growth, Growth Measures, and Standard Errors



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

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

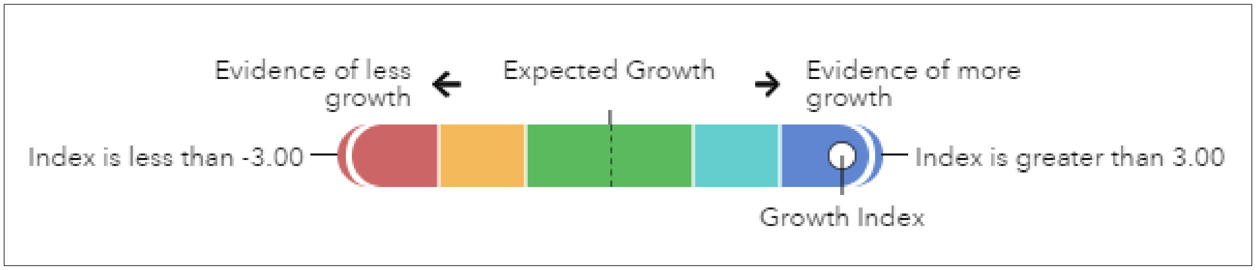
$$Growth\ Index = \frac{Growth\ Measure - Expected\ Growth}{Standard\ Error\ of\ the\ Growth\ Measure} \tag{23}$$

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

- **Well Above** indicates significant evidence that students made more growth than expected. The growth index is 2 or greater.
- **Above** indicates moderate evidence that students made more growth than expected. The growth index is between 1 and 2.
- **Meets** indicates evidence that students made growth as expected. The growth index is between -1 and 1.
- **Below** indicates moderate evidence that students made less growth than expected. The growth index is between -2 and -1.
- **Well Below** indicates significant evidence that students made less growth than expected. The growth index is less than -2.

This is represented in the growth indicator bar in Figure 9, which is similar to what is provided in the Division and School Value-Added reports in the VVAAS 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 6: Sample Growth Indicator Bar



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

4.5 Rounding and Truncating Rules

As described in the previous section, the definitions of the growth categories are 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.

5 Input Data Used in VVAAS Growth Model

5.1 Assessment Data

For the analysis and reporting based on the 2021-22 school year, EVAAS receives the following assessments for use in the growth and/or projection models:

- SOL Mathematics and Reading in grades 3–8
- SOL Science in grades 5 and 8
- SOL Writing in grade 8
- SOL Civics and Economics
- SOL VA Studies
- SOL EOC assessments in Algebra I, Algebra II, Biology, Earth Science, Geometry, Reading, Virginia History, World Geography, World History I, World History II, Writing

While not used in the growth or projection models, the following assessments are included in web reports related to students' testing history:

- PALS Literacy K-3
- SOL EOC Chemistry and United States History
- VGLA Reading in grades 3-8 (2016 and 2017 only)
- VAAP Mathematics in grades 3-12 and Reading in grades 3-8
- VAAP Science in grades 5-12
- VAAP History in grades 4-12
- VAAP Writing in grade 8
- VAAP EOC Reading in grades 9-12
- VAAP EOC Writing in grades 9-12

Assessment files provide the following data for each student score:

- Student identifiers
  - Student First Name
  - Student Last Name
  - Student Middle Initial
  - State Testing ID
  - Student date of birth
- Scale Score
- Performance Level
- Subject
- Tested Grade
- Tested Semester
- Division Number
- School Number
- Membership
  - School (Been Enrolled in School)
  - Division (Not Enrolled in School but Enrolled in Division)

- State (Not Enrolled in Division but Enrolled in a Virginia Public Division)
- Not in VA (Not Enrolled in a Virginia Public Division)

- Testing Status
  - Nullified
  - Medically Exempt
  - Did Not Attempt
  - Absent
- Test Form/Version/Modified Test Format
  - Large Print
  - Braille
  - ELSA

5.2 Student Information

Student information is used in creating the web application to assist educators, analyze the data to inform practice, and assist all students with academic growth. SAS receives this information in the form of various socioeconomic, demographic, and programmatic identifiers provided by VDOE. Currently, these categories are as follows:

- Gender (M, F, N)
- Race
  - American Indian or Alaska Native
  - Asian
  - Black, not of Hispanic Origin
  - Hispanic
  - Native Hawaiian/Other Pacific Islander
  - Non-Hispanic, two or more races
  - White, not of Hispanic Origin
- Students with Disabilities (Y, N)
- Economically Disadvantaged (Y, N) – only reported at the aggregate level
- English Learner (Y, N)

6 Business Rules

6.1 Assessment Verification for Use in Growth Models

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

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

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

6.1.1 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 Virginia assessments are well below acceptable values, meaning that these assessments have adequate stretch to measure growth even in situations where the group of students are very high or low achieving.

6.1.2 Relevance

Relevance indicates whether the test is sufficiently aligned with the curriculum. The requirement that tested material correlates with standards will be met if the assessments are designed to assess what students are expected to know and be able to do at each grade level. More information can be found at the following link: https://www.doe.virginia.gov/teaching-learning-assessment/student-assessment/virginia-sol-assessment-program

6.1.3 Reliability

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

6.2 Pre-Analytic Processing

6.2.1 Missing Grade

In Virginia, the grade used in the analyses and reporting is the tested grade, not the enrolled grade, for grade-level assessments such as SOL Math and Reading in grades 3-8 and SOL Science in grades 5 and 8. If a grade is missing on a grade-level test record (meaning grades 3-8), then that record will be excluded from all analyses. The grade is required to include a student's score in the appropriate part of the models for those assessments.

6.2.2 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 record that has more demographic fields filled out will be retained for the analysis and reporting and the other records will be excluded from the analysis and reporting.

6.2.3 Students with Missing Divisions or Schools for Some Scores but Not Others

If a student has a score with a missing division or school for a particular subject and grade in a given testing period, then the duplicate score that has a division and/or school will be included over the score that has the missing data. This rule applies individually to specific subject/grade/years.

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

If a student has multiple scores in the same period for a particular subject and grade and the test scores are not the same, then the model will retain any proficient score over a non-proficient score and use the earliest test date with the highest score.

6.2.5 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 all records with multiple grades are retained in the analysis dataset.

6.2.6 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.

6.2.7 Outliers

Student assessment scores are checked each year to determine whether they are outliers in context with all the other scores in a reference group of scores from the individual student. These reference scores are weighted differently depending on proximity in time to the score in question. Scores are checked for outliers using related subjects as the reference group. For example, when searching for outliers for Math test scores, all SOL and SOL EOC Math subjects are examined simultaneously, and any scores that appear inconsistent, given the other scores for the student, are flagged.

Scores are flagged in a conservative way to avoid excluding any student scores that should not be excluded. Scores can be flagged as either high or low outliers. Once an outlier is discovered, then that outlier will not be used in the analysis, but it will be displayed on the student testing history on the VVAAS 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 for SOL Math and Reading when determining the difference between the score in question and the weighted combination of reference scores (otherwise known as the comparison score). In other words, the score in question must be at least 3.5 standard deviations below the comparison score. For other assessments, the t-value must be below -4.0.
- The percentile of the comparison score must be above a certain value. This value depends on the position of the individual score in question but will range from 10 to 90 with the ranges of the individual percentile score.

For high-end outliers, the rules are:

- The percentile of the score must be above 50.
- The t-value must be above 4.5 for SOL Math and Reading when determining the difference between the score in question and the reference group of scores. In other words, the score in question must be at least 4.5 standard deviations above the comparison score. For other assessments, 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.

6.2.8 Linking Records over Time

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

6.3 Growth Models

6.3.1 Students Included in the Analysis

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

For the predictive and projection models, a student must have at least two valid predictor scores that can be used in the analysis, all of which cannot be deemed outliers (see Section 6.2.7 on Outliers). These scores can be from any year, subject, and grade that are used in the analysis. In other words, the student's expected score can incorporate other subjects beyond the subject of the assessment being used to measure growth. The required two 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 two-score minimum, then that student is excluded from the analyses. It is important to note that not all students have to have the same two prior test scores; they only have to have some subset of two that were used in the analysis. This is because the predictive model does not determine growth based on consecutive grade movement on tests, and as such, students do not need to stay in one cohort from one year to the next. That said, if a student is retained, retakes the same test, and passes that test, then the student's retake score will not be included in the model while the student's original test score is included. 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 model, it is typically the case that a prior test is only considered a possible predictor when at least 50% of the students used in that model have those prior test scores.

6.3.2 Minimum Number of Students to Receive a Report

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

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