SAS® EVAAS

Statistical Models and Business Rules

Prepared for the Center for the School of the Future at Utah State University



Contents

1 Introduction to Utah's Value-Added Reporting 2

2 Statistical Model 3
2.1 Overview of Statistical Model 3
2.2 Technical Description of the Predictive Model 3
2.3 Calculation of Percentage of Students Exceeding Predicted Achievement 4
2.4 Outputs from the Models 5

3 Input Data Used in EVAAS Growth Model 6
3.1 Assessment Data 6
3.2 Student Information 6
3.3 Student-Teacher Linkages 7

4 Business Rules 8
4.1 Assessment Verification for Use in Growth Models 8
4.2 Student-Teacher Linkages and Courses 9
4.3 Student-Teacher Linkages and Grade Level 9
4.4 Selection of Acadience Scores Among Multiple Scores in the Same Testing Window 9
4.5 Outliers 9
4.6 Linking Records Over Time 10

5 Growth Models 11
5.1 Instructional Responsibility for Students Claimed by Multiple Teachers 11
5.2 Participation Code Exclusions from Models as Outcomes and Predictors 11
5.3 Participation Code Exclusions from Models as Outcomes / Inclusion as Predictors 11
5.4 Attendance Exclusion 12
5.5 Enrollment Exclusion 12
5.6 Minimum Number of Prior Scores (Predictors) 12
5.7 Minimum Number of Students for Teacher Composite 12

100126114-569. Copyright © 2026 SAS Institute Inc. Cary, NC, USA. All rights reserved. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

1 Introduction to Utah's Value-Added Reporting

SAS® EVAAS supports the implementation of the Teacher Merit Award (TMA) program by providing analytic support and calculating value-added results for use in the program. These value-added results use EVAAS models.

EVAAS models use established statistical analyses to measure students' academic growth. Conceptually, value-added measures are calculated by comparing the prior achievement to the current achievement for a group of students. Although this concept is easy to understand, the implementation of a value-added model is more complex. This document provides details about the business rules used for calculating value-added measures for use in TMA.

In order to provide reliable value-added measures, models must overcome the complexities of working with student assessment data. A few examples are students who do not have the same entering achievement, students who do not have the same set of prior test scores, and all assessments have measurement error because they are estimates of student knowledge or skill.

Examining the difference between a student's actual achievement and their predicted achievement creates a measure that is more nuanced and statistically robust.

With these complexities in mind, the purpose of this document is to provide information about the EVAAS model used to support TMA. More model information in these areas is provided throughout this document:

- Conceptual and technical explanations of analytic models
- Definition and calculation of predicted achievement
- Determination of the percentage of students exceeding predicted achievement
- Input data
- Business rules

2 Statistical Model

2.1 Overview of Statistical Model

The predictive model is a regression-based value-added model where growth is analyzed by comparing students' predicted achievement and their actual scores. For each assessment, grade, and year, the model uses qualifying statewide Utah student records to estimate the statistical relationship between current achievement and eligible prior test scores. It then applies those statewide-estimated relationships to the individual student's own available prior-score history, weighting the prior scores according to their joint relationship with the current assessment. Thus, statewide data supports the prediction equation, while the student's own prior achievement individualizes the expected score. A student exceeds predicted achievement when the observed current-year score is higher than this individualized expected score that represents predicted achievement. This model is used for the purposes of the Teacher Merit Award (TMA) program.

2.2 Technical Description of the Predictive Model

Each student receives a scale score representing their predicted achievement based on their own prior testing history and the predictive model using qualifying statewide Utah student records to estimate the statistical relationship between current achievement and eligible prior test scores. In practical terms, this score represents the student's entering achievement because it is based on all prior testing information to date. The scale scores representing students' predicted achievement can be compared to the students' actual scores to determine whether students have exceeded predicted achievement.

The approach is described briefly below with more details following.

- The predicted achievement score serves as the response variable (y, the dependent variable).
- The covariates (x terms, predictor variables, explanatory variables, independent variables) are scores on tests the student has already taken.

Algebraically, the model can be represented as follows for the ith student.

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

The μ terms are means for the response and the predictor variables. 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 actual value for the specific response and have three predictor scores. The resulting prediction equation for the ith student is as follows:

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

One difficulty that must be addressed to implement the predictive model is that not all students will have the same set of predictor variables due to missing test scores. The strategy for dealing

with missing predictors is to estimate the joint covariance matrix (call it C) of the response and the predictors. Let C be partitioned into response (y) and predictor (x) partitions, that is,

$$C = \begin{bmatrix} c_{yy} & c_{yx} \\ c_{xy} & C_{xx} \end{bmatrix} \quad (3)$$

This matrix is estimated using the Expectation Maximization (EM) algorithm for estimating covariance matrices in the presence of missing data provided by the Multiple Imputation procedure in SAS/STAT® (although no imputation is actually used). Only students who had a test score for the response variable in the most recent year and who had the required number of variables are included in the estimation. Given such a matrix, the vector of estimated regression coefficients for the prediction equation (2) can be obtained as:

$$\hat{\beta} = C_{xx}^{-1} c_{xy} \quad (4)$$

This allows the use of whichever predictors a student has to get that student's predicted y-value ($\hat{y}_i$). Specifically, the C_{xx} matrix used to obtain the regression coefficients *for a particular student* is a subset of the overall C matrix that corresponds to the set of predictors for which this student has scores.

The prediction equation also requires estimated mean scores for the response and for each predictor (the $\hat{\mu}$ terms in the prediction equation). Once the parameter estimates for the prediction equation have been obtained, predictions can be made for any student with any set of predictor values, so long as that student meets the minimum requirement for number of prior test scores.

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

The $\hat{y}_i$ term is nothing more than a composite of all the student's past scores. It is a one-number summary of the student's level of achievement prior to the current year, and this is sometimes referred to as the expected score, entering score, or predicted achievement. The different prior test scores making up this composite are given different weights (by the regression coefficients, the $\hat{\beta}$ terms) to maximize its correlation with the response variable. Thus, a different composite would be used when the response variable is Math than when it is Reading, for example. $\hat{y}_i$ represents prior achievement before the effect of the current district, school, or teacher.

2.3 Calculation of Percentage of Students Exceeding Predicted Achievement

For the purposes of TMA, the $\hat{y}$ terms, expressed in scale scores, are compared to the observed scale score for each student for each assessment to determine whether they have exceeded their predicted achievement. These are then aggregated to the teacher level for each applicable assessment for teachers using teacher-student linkages. These aggregations are produced for each assessment for each teacher, and all students across all grades and subjects for which the teacher has results are aggregated into a single composite. The composite value is used for the purposes of TMA. This percentage is calculated as follows:

$$\%MEE = \frac{\sum_{i=1}^{n}(w_i S_i)}{\sum_{i=1}^{n} w_i}, where\ S_i = \begin{cases} if\ y_i > \hat{y}_i \\ if\ y_i \leq \hat{y}_i \end{cases} \quad (6)$$

When calculating this percentage, a teacher's percentage of instructional responsibility, or w_i, for student *i* is used to weight the results. If different students are claimed at different percentages of instructional responsibility for a given teacher, each is weighted proportionally when calculating the percentage of students exceeding predicted achievement for that teacher. In equation 6, S_i = 1 if the observed scale score (y_i) for that student is greater than the predicted achievement for that student, and S_i = 0 if the observed scale score (y_i) for that student is less than or equal to the predicted achievement for that student. These values are then multiplied by the instructional responsibility percentage in the numerator, and that student's contribution to the denominator is the instructional responsibility percentage. This is done for all students for a given assessment and for the composite that goes across all assessments for which the teacher has results. The resulting numerator and denominator include all applicable students to calculate a percentage of students meeting or exceeding expectations.

For example, if a teacher has 30 students that all exceeded their predicted achievement and were all claimed at 100 percent instructional responsibility, each student contributes values of 1 to both the numerator and denominator, resulting in 30 divided by 30, or 100% of students exceeding predicted achievement. If the same teacher had 50 percent instructional responsibility for the same set of students, the numerator would be 15 and the denominator would be 15, also resulting in 100% of students exceeding predicted achievement.

Alternatively, in some cases some students for a teacher might be claimed at 50 percent instructional responsibility and others might be claimed at 100 percent instructional responsibility. The students with 50% instructional responsibility for the teacher contribute 0.5 to the numerator and denominator if they exceed predicted achievement or 0 to the numerator and 0.5 to the denominator if their observed scale score is less than or equal to predicted achievement. The students with 100% instructional responsibility for the teacher contribute 1 to the numerator and denominator if they exceed predicted achievement or 0 to the numerator and 1 to the denominator if their observed scale score is less than or equal to predicted achievement. In practice, this means that students that have higher percentages of instructional responsibility have a proportionally larger contribution to a teacher's percentage of students exceeding expectations than those that have lower percentages of instructional responsibility.

2.4 Outputs from the Models

Results are provided for the following assessments:

- Acadience Early Literacy Grades K–3
- Acadience Early Numeracy Grades K–3
- RISE English Language Arts Grades 3–8
- RISE Mathematics Grades 3–8
- RISE Science Grades 4–8
- Utah Aspire Plus Math Grades 9 and 10
- Utah Aspire Plus Language Arts Grades 9 and 10

3 Input Data Used in EVAAS Growth Model

3.1 Assessment Data

To support analysis and supporting, EVAAS receives assessment data for use in the value-added models:

- Acadience Early Literacy Grades K–3
- Acadience Early Numeracy Grades K–3
- RISE English Language Arts Grades 3–8
- RISE Mathematics Grades 3–8
- RISE Science Grades 4–8
- Utah Aspire Plus Math Grades 9 and 10
- Utah Aspire Plus Language Arts Grades 9 and 10

Assessment files provide the following data for each student score:

- USBE Student ID
- Test Taken
- Tested Subject
- Tested Grade
- Tested Period (Received only for Acadience)
- District Number
- School Number
- Scale Score
- Performance Level (Received only for RISE and Utah Aspire Plus)
- Participation Code (Received only for RISE and Utah Aspire Plus)

3.2 Student Information

Additional student information is received to support the analysis and the reporting available on Utah Learns where applicable. Currently, this data includes:

- USBE Student ID
- Student First Name
- Student Last Name
- English Language Learner (Yes, No)
- Low Income (Yes, No)
- Race/Ethnicity (American Indian, AfAm/Black, Asian, Hispanic, Multiple Race, Pacific Islander, White)
- Sex (Male, Female)
- Special Education (Yes, No)
- Excused Absences
- Unexcused Absences
- Attendance
- Enrollment Entry Date
- Enrollment Exit Date

3.3 Student-Teacher Linkages

The Center for the School of the Future via a data sharing agreement with the Utah State Board of Education provided SAS EVAAS with student, teacher, and course data that is used to establish linkages for the purpose of the analysis. Where applicable, this data includes:

- Teacher Identifiers (Teacher CACTUS ID, Teacher First Name, Teacher Last Name)
- Student Identifiers (USBE Student ID, Student First Name, Student Last Name)
- Course information for identification of courses associated with rostered assessments
- District Number
- School Number

4 Business Rules

4.1 Assessment Verification for Use in Growth Models

To be used appropriately in any growth models, assessments must meet four criteria. The first two are considered by assessment developers and psychometricians who support the implementation of the assessments and represent characteristics of high-quality assessments. The last two are reviewed annually by the SAS EVAAS team to confirm each assessment's suitability for use in growth modeling:

1. The test is **highly relevant** to the academic standards of the course it is designed to assess so that it is possible to measure progress with the assessment in that subject, grade, and year.
2. The scales are **reliable**, meaning the assessment is of high quality and constructed using psychometric standards and properties.
3. 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.
4. The observed relationships between available prior scores and the outcomes for a given assessment in a given year show **sufficient predictability**, demonstrating that the statistical relationships are strong enough to support modeling.

These criteria should be met for any assessment used in any growth model.

The first two criteria can be considered characteristics of high-quality assessments. Assessment developers and psychometricians ensure that the assessment is aligned with the academic standards the assessment is meant to assess. 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. In addition to this relevance of the content of the assessment, high-quality assessments are designed to produce reliable results. These criteria are typically checked and monitored by assessment developers and psychometricians that support the ongoing implementation of the assessments.

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 assessments are well below acceptable values, meaning that these assessments have adequate stretch to measure value-added even in situations where the group of students are very high or low achieving. The SAS EVAAS team checks the stretch of each assessment in each year to confirm suitability for use in growth modeling.

The SAS EVAAS team also monitors each assessment in each year to confirm that the available prior scores are sufficiently related to the current year assessment outcomes. This includes reviewing the correlations between each available prior score with each outcome score and reviewing the ability of all prior scores collectively to predict each outcome score.

4.2 Student-Teacher Linkages and Courses

Core code mappings are used to identify in-scope core codes (courses) and the associated test/subject. The test and subject are assigned as indicated by the publicly posted USBE mapping. The test and subject are used to connect the student-teacher linkages to assessment data.

4.3 Student-Teacher Linkages and Grade Level

For Acadience and RISE assessments, student-teacher linkages are connected to student assessments regardless of student assessment/course grades. This means that a student that is enrolled in a grade 4 course but takes a grade 5 assessment would be connected to the teacher for these purposes. For Utah Aspire Plus, the grade level of the assessment and the grade level of the course data must match to be included; specifically, both must be from grade 9 or both must be from grade 10.

4.4 Selection of Acadience Scores Among Multiple Scores in the Same Testing Window

For Acadience assessments, testing windows are provided in the assessment data but test dates are not included. In the Fall and Winter windows, the assessment with the lowest score is used to ensure there is only one record per subject and grade within a testing window. For Spring assessments, the highest score is used.

4.5 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 Math assessments 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.

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), except in the case of RISE Science and Utah Aspire Plus Language Arts and Math, where the t-value must be below -4.0. In other words, the score in question must be at least 3.5 (or 4) standard deviations below the comparison score.
- 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 when determining the difference between the score in question and the reference group of scores, except for RISE Science and Utah Aspire Plus Language Arts and Math, where the t-value must be above 5.0. In other words, the score in question must be at least 4.5 standard deviations above the comparison score. For Science 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.

4.6 Linking Records Over Time

Student assessment 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 the USBE Student ID. Teacher records are matched over time using the teacher CACTUS ID.

5 Growth Models

5.1 Instructional Responsibility for Students Claimed by Multiple Teachers

In the 2024-25 analysis, all student-teacher linkages are considered 100% instructional responsibility regardless of whether multiple teachers are connected to a student for a given course or assessment. In the 2025-26 analysis, percentages of instructional responsibility reported during Roster Verification will be used where applicable.

5.2 Participation Code Exclusions from Models as Outcomes and Predictors

For RISE and Utah Aspire Plus assessments, records with the following participation code values are not included in the model when calculating results or as predictors in models for subsequent assessments:

- 101 - Did Not Test
- 106 - Student Refused to Test
- 107 - Excused for Health Emergency
- 108 - Course Will Continue Next School Year
- 109 - Course Instruction Not Aligned with Course Code
- 111 - USBE Excused - Approval Needed
- 112 - Student Transferred Before Testing
- 202 - Modified
- 203 - Invalidated
- 204 - Parental Exclusion
- 208 - Test System Irregularity
- 209 - Incorrect Course Code Assigned
- 213 - Invalidated due to Academic Integrity – Student
- 214 - Invalidated due to Parental
- 301 - USBE Assigned Did Not Test
- 303 - USBE Assigned Invalidated

5.3 Participation Code Exclusions from Models as Outcomes / Inclusion as Predictors

For RISE and Utah Aspire Plus assessments, records with the following participation codes values will not receive predicted achievement scores and thus will be excluded from the results, though in contrast with the codes listed in 4.3.2, past scores with these codes *will* be included when predicting achievement on a subsequent assessment:

- 103 - EL First Year in U.S. April 15 or Later
- 104 - EL First Year in U.S. Before April 15
- 313 - EL First Year in U.S. April 15 or Later Student tested
- 314 - EL First Year in U.S. April 15 or Later

5.4 Attendance Exclusion

Students are excluded from the results for a teacher if they were absent for at least 10% of days at the school for which the student is connected to the teacher. The percentage of absences is calculated by dividing the total number of absences by the sum of the total number of absences and days noted in attendance. All records received for a student for a given school are considered together.

This approach aligns with the rules specified in Utah Code § 53G-11-520 and Utah Administrative Rule R277-323-3(6).

5.5 Enrollment Exclusion

Students are excluded from the results for a teacher if they were enrolled at the school for which the student is connected to the teacher for fewer than 60 calendar days. If a student has enrollment records across multiple schools, this rule is applied separately for each school based on their length of enrollment there. The length of enrollment is calculated using the entry date and exit date, and if an exit date is missing the last day of school provided by USBE is used. This rule is specified in the Utah Code § 53G-11-520 and Utah Administrative Rule R277-323-1(4).

5.6 Minimum Number of Prior Scores (Predictors)

For most assessments, the analysis excludes scores for students who do not have at least three prior test scores (predictors) from the predictor list. The exceptions are:

- For Acadience Kindergarten, only one predictor is required. The Beginning of Year (BOY) scores are not included as predictors in the model due to low differentiation. Additionally, if only one prior test score is used to make a prediction, the prior test score must be of the same subject as the predicted score and taken during the middle-of-year (MOY) period.
- For RISE Grade 4, only two predictors are required.

5.7 Minimum Number of Students for Teacher Composite

A teacher must have at least four students meeting all criteria for inclusion in the modeling to receive an across grades/assessments composite for the purposes of use in TMA. These four students can have a combination of grades/assessments, and a student counts toward the total for each assessment for which they are linked to the teacher and are included in the analysis.