

SAS[®] EVAAS

URM Modeling Approach for Value-Added

Introduction

This document provides an overview of the value-added modeling approach used for OST end-of-course (EOC), Science, and Social Studies tests. Tests that are not given for consecutive years require a different modeling approach from the gain-based approach (known as the MRM) used for OST Math and English Language Arts (ELA). To be more specific, although some EOC tests are taken in succession, not all students take them in the same grades, and this approach is more appropriate in those scenarios.

For Ohio's value-added reporting, the modeling approach for OST EOC, Science and Social Studies tests is called the *univariate response model* (URM). The URM is a regression-based model, which measures the difference between students' expected scores for a subject/year with their actual scores. Students' expected scores are based on the relationships between the test in question and prior test scores, meaning there must be a sufficient relationship between these scores to provide an expected score. The growth expectation is met when, on average, students linked to a specific district/school/teacher made the same amount of growth as students linked to the average district/school/teacher with the state for that same year/subject/grade.

Key Technical Aspects of the URM Approach

The URM approach shares many of the statistical advantages as the MRM approach. The URM uses all available testing data for each student, accommodates teaching scenarios where more than one teacher has responsibility for a student's learning in a specific subject/grade/year, and does not require all test scores to be on the same scale. Although some value-added/growth models require students to have all predictors or the same set of predictors, the URM does not. Students can be included if they have at least three prior test scores in any subject/grade. As mentioned above, the URM does not require consecutive grade given tests while the MRM does.

Common Questions for Interpretation

What is student growth? Student growth looks at how much academic growth students make within a tested subject/grade over time.

What is the modeling approach? For tests given in non-consecutive grades, growth is measured through a *regression-based approach*, which compares students' expected scores with their actual scores. A student's expected score is obtained by looking at their prior testing performance (scores) across all subjects, as well as all students with similar prior testing performance, and estimating how that group of students scored, on average, on that test.

How is a student's expected score determined? The process of generating each student's expected score begins with building a robust statistical model of all students who took the selected assessment in the most recent year. The model includes the scores of all students in the state or reference group along with their testing histories across years, grades, and subjects. This model determines the relationships of all the prior assessment scores to students' performance on the selected assessment.

This model determines the relationships between all students' prior assessment scores and their actual performances on the selected assessment. By considering how all other students performed on the assessment in relation to their testing histories, the model is able to calculate an expected score for each student based on their individual testing history.

The model uses prior test scores from all grades and subjects as this improves the reliability in determining the expected score. (For example, eighth-grade Math, ELA, and Science scores are typically related to the Algebra I scores.) Furthermore, the prior test scores are weighted according to the strength of the relationship that they have with the end-of-course subject, so Math scores can be weighted more than ELA scores for the Algebra I analysis although ELA scores can still help improve the prediction.

Conceptually, it might be useful to consider the expectation as the student's entering achievement since it is based on all the student's achievement data prior to the year in question.

What is the growth measure? The growth measure is related to the number of scale score points a group of students scored *above* or *below* their expected score, which considers their prior testing performance. The growth measure is not an exact difference between the expected and actual scores for a group of students. To provide additional protections to educators, the growth measure uses shrinkage estimation, which means that each district/school/teacher is assumed to have the average growth measure (zero) until the weight of the evidence pulls the growth measure above or below zero. In other words, the growth measure is a function of the difference between the expected and actual scores for a group of students. This function will consider the number of students used in the growth measure as well as the variability of the districts/schools/teachers across that entire subject and grade for a given year. The more students a teacher has, the more certainty there is on students' growth, and this means less shrinkage. When less variability among school or teacher growth measures is observed throughout the state, more shrinkage will be applied to those school or teacher growth measures. Finally, at the teacher level, this function also accounts for the sharing of instructional responsibility.

How is the growth measure calculated? The key steps are as follows:

- Each student's prior testing performance is used to determine their expected score.
- The expected score assumes that each student has the average schooling experience for the reference population.
- The expected score is compared to the actual score for each student.
- The growth measure is a function of the difference of the expected and actual scores for a group of students linked to a teacher, school, or district.

What is the standard error of the growth measure? This is a measure of certainty of the estimated school or teacher growth measure, and it provides insight regarding the certainty that a growth measure is decidedly above or below expected growth, protecting educators and schools from misclassification. The standard error is based on both the quantity and quality of data for the school or teacher, such as the number of students linked to the teacher or school and the completeness of the students' test scores. Note, standard error relates to certainty around a teacher's individual growth measure and is different for each teacher. Therefore, the standard error is not forcing any preset percentages of teachers into the different categories. In other words, the standard error is not the standard deviation of the teacher value-added measures.

What is the growth index? The growth index is then calculated by dividing the growth measure by its standard error in order to determine the significance of the growth measure.

What is the “the average schooling experience”? For School Value-Added reports, the average schooling experience means the growth that students made in the average school from the reference population for that course in the current year. In other words, for Algebra I School Value-Added reports, the average schooling experience represents the progress that Algebra I students made in the average school in the reference population. This reference group is based on the pool of test-takers, which is statewide for state assessments. The average schooling experience considers each student’s prior testing performance.

For Teacher Value-Added reports, the average schooling experience means the growth that students made who were linked to the average teacher from the reference population for that course in the current year. This ensures an “apples-to-apples” comparison among schools and teachers.

What is the expected growth and how is it calculated? Expected growth means that the actual scores of a group of students linked to a teacher/school/district were, on average, the same as their expected scores. In other words, the difference between students’ expected and actual scores was zero (or close to zero).

It is possible to have a small negative effect or small positive effect yet still be considered “making expected growth.” Because the growth measure is an estimate of student growth, there must be sufficient evidence that a growth estimate is decidedly above or below expected growth to receive the green or red labels.

At a Glance - Understanding the URM Value-Added Report

At a glance, you can see whether students with a teacher, school, or district made more, less, or about the expected growth due to the color-coding. These colors can be interpreted as follows:

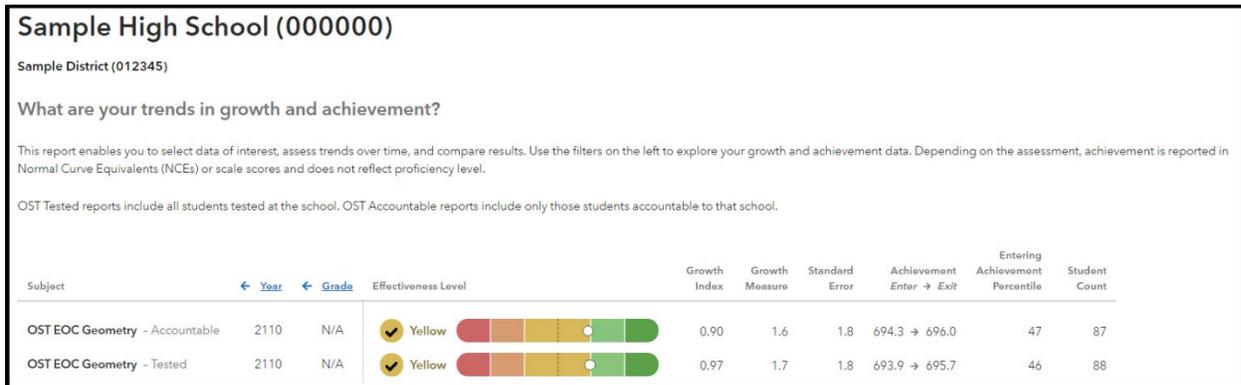
-  Dark Green *Significant* evidence that students made more growth than expected.
-  Light Green *Moderate* evidence that students made more growth than expected.
-  Yellow Evidence that students made growth as expected.
-  Orange *Moderate* evidence that students made less growth than expected.
-  Red *Significant* evidence that students made less growth than expected.

This information provides educators with a quick indicator of whether they are meeting the academic needs of all their students. Other EVAAS reports (Diagnostic reports, Decision Dashboard, etc.) can provide insight as to why this is the case.

Deeper Dive - Understanding the URM Value-Added Report

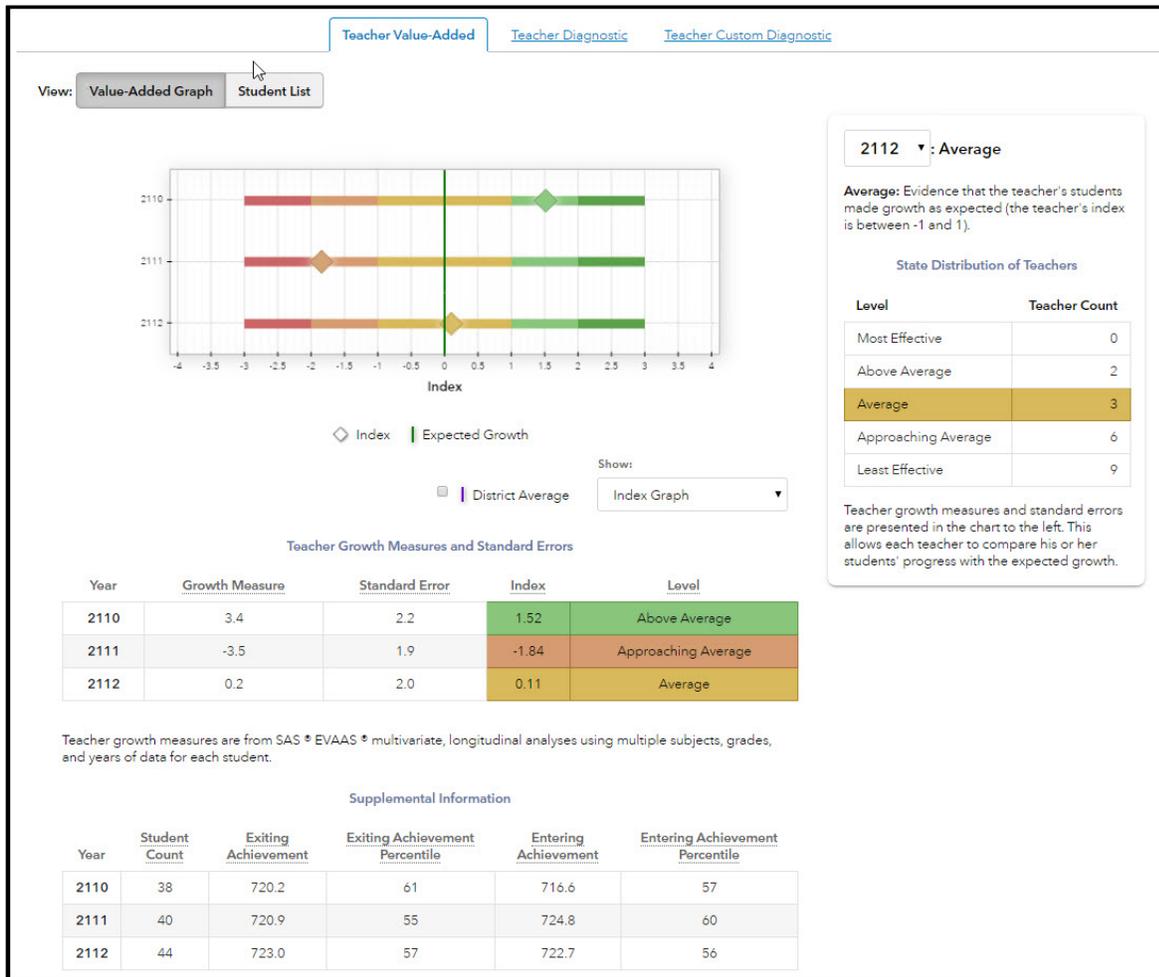
The Value-Added reports are available through the EVAAS web application. The reports are designed to provide basic information about the analysis (number of students included, students’ entering achievement, etc.) as well as to provide quick feedback on students’ growth over time.

Figure 1: Sample URM School Value-Added Report



- **Effectiveness Level** is a category that describes the certainty that a group of students met, exceeded, or fell short of expected growth.
- **Growth Index** is an indicator of certainty that the group of students met, exceeded, or fell short of expected growth.
- **Growth Measure** is a conservative estimate of the growth that students made, on average, in a grade and subject or course.
- **Standard Error** is a measurement that establishes a confidence band around the growth measure and describes the certainty that the group of students met, exceeded, or fell short of expected growth.
- **Achievement** is reported in scale scores.
 - Entering achievement is the average of the students' expected scale scores.
 - Exiting achievement is the average of the students' actual scale scores.
- **Entering Achievement Percentile** is the entering achievement for the group of students relative to the overall distribution for this assessment.
- **Student Count** is the number of students included in the analysis.

Figure 2: Sample URM Teacher Value-Added Report



- **Growth Measure** is a conservative estimate of the growth that students made, on average, in a grade and subject or course.
- **Standard Error** is a measurement that establishes a confidence band around the growth measure and describes the certainty that the group of students met, exceeded, or fell short of expected growth.
- **Index** is an indicator of certainty that the group of students met, exceeded, or fell short of expected growth.
- **Level** is a category that describes the certainty that a group of students met, exceeded, or fell short of expected growth.
- **Student Count** is the number of students included in the analysis.
- **Exiting Achievement** is the average of the students' actual scale scores.
- **Exiting Achievement Percentile** is a simple average of scores for the students included in the analysis for each year of data. This value is expressed in scale score points.
- **Entering Achievement** is the average of the students' expected scale scores.

- **Entering Achievement Percentile** is the entering achievement for the group of students relative to the overall distribution for this assessment.