SAS® EVAAS

How will EVAAS accommodate changes in Ohio's tests?

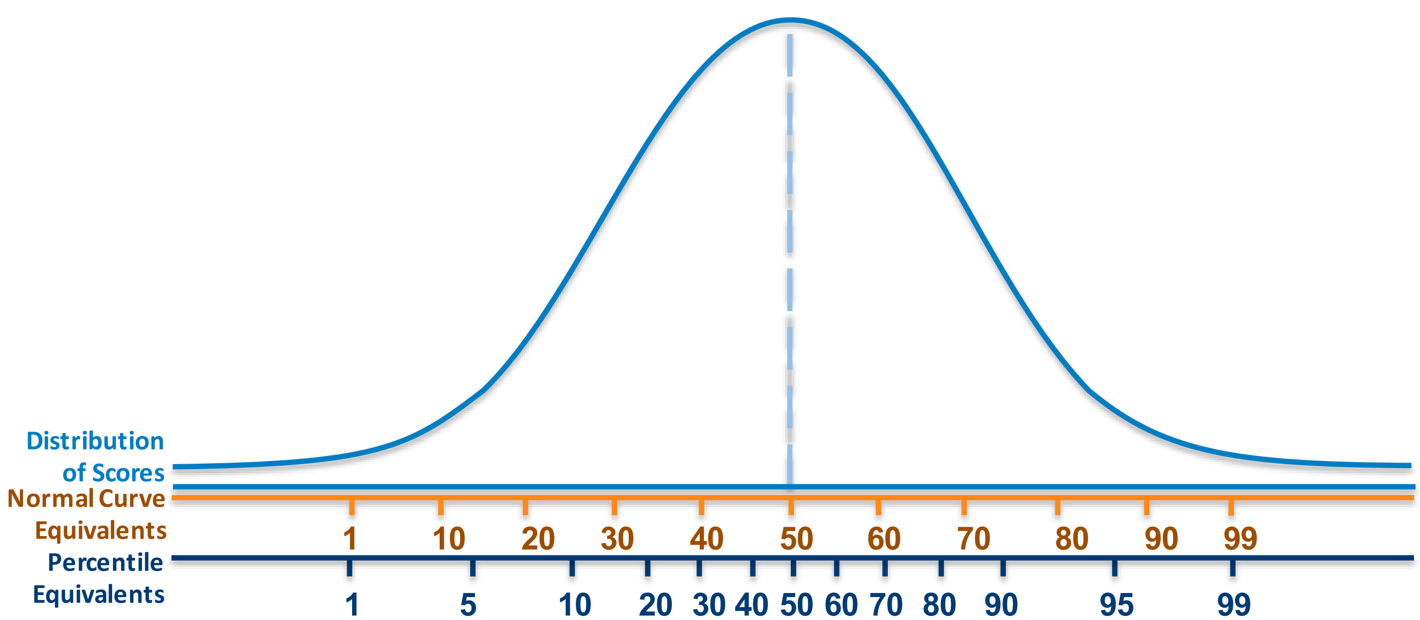
Introduction

EVAAS value-added analyses have been available to educators and policymakers for the past two decades. During that time, there have been many changes in testing regimes in the states that use EVAAS. From this experience, SAS has developed a statistically robust process to use all available testing data from both old and new tests as well as to ensure continuity in reporting. By using some simplified examples of EVAAS models, this document provides an overview of how EVAAS will accommodate changes in Ohio's tests and continue to provide valid and statistically robust value-added results.

How do EVAAS models typically calculate value-added measures?

EVAAS value-added measures are calculated using all available student test scores across all grades and subjects for the most recent five years. The multivariate response model (MRM) is a gain-based approach, and it is used for OST Math and Reading in grades 4–8. To measure gains from one grade to the next, the student scale scores are first converted to a Normal Curve Equivalent (NCE) distribution based on the state population for a particular test/subject/grade/year. The NCE distribution matches a percentile distribution at 1, 50, and 99. However, unlike a percentile distribution, the NCE distribution is equal-interval such that the difference between 50 and 60 is the same as the difference between 80 and 90. This is an important distinction for measuring gains to ensure that educators are neither advantaged nor disadvantaged by the entering achievement of their students. See Figure 1 for a graphical representation of these distributions.

Figure 1: Sample Distribution of Scores, NCEs, and Percentiles



To simplify the MRM approach, the average NCE gain for students with a district, school, or teacher is based on their change in achievement as measured in NCEs. For example, if students in a school entered at the 50th NCE and left at the 55th NCE, then the value-added measure would be the difference between 55 and 50, or 5 NCEs. If students entered at the 50th NCE and left at the 50th NCE, then their gain would be zero. However, this does not represent zero growth; it simply means that students have maintained their position in the distribution. Typically, zero represents the expected growth for students.

However, setting the growth expectation is a bit more complicated. Through 2014, the NCE distribution was calculated in terms of the 2010 (“base year”) distribution of scores in a particular subject/grade for OAA Math and Reading. The NCE distribution can also be calculated in terms of the current year – creating an “intra-year” computation. That approach has been used in Ohio since 2015. These two approaches, a “base year” or an “intra-year,” are described briefly below.

- **Base year:** a cohort of students from grade to grade maintains the same relative position with respect to the statewide student achievement in the base year for a specific subject and grade.
- **Intra-year:** students maintained the same relative position with respect to the statewide student achievement that year.

In the base year approach, the student scale scores are translated into NCEs based on the distribution of scores from the base year, and the gains are thus relative to the achievement in the different grades of the base year distribution. The base year approach assumes that scaling is stable from one year to the next, meaning in a specific subject and grade a scale score must mean the same thing from one year to the next. Thus, any transition to a new test would not uphold that assumption.

When an intra-year approach is used, the base year is essentially re-set each year so that gains are considered only in terms of their relative position each year. Because the gains are relative, it is not necessary for the tests to be in the same scale in measuring gains from one year to the next.

Again, in either approach, an average NCE gain is calculated for a district, school, or teacher in a given test/subject/grade/year as a value-added measure. This value is compared to an expected growth of zero, which represents that this group of students, on average, maintained the same relative position in the statewide distribution from one grade to the next in a given subject.

The MRM approach is typically used for tests given in consecutive grades, like OST Math and Reading. For tests that are given in non-consecutive grades, such as OST Science, the univariate response model (URM) can be used to measure growth. Rather than a gain-based model, the URM is a regression-based model where the growth measure is based on the difference between students’ expected achievement and their actual achievement. The expected achievement is based on students’ prior testing histories, and these do not have to be on the same scale. The expectation of progress is based on how students in the reference population did in the current year. In this way, the URM can be thought of as somewhat similar to the intra-year approach. In other words, there is not a comparison to a base year; the growth expectation is determined within the current year. The growth expectation of the URM can be described as follows:

- **Intra-year:** students in a district/school/teacher’s class made the same amount of progress as students in the average district/school/teacher in the state for that same year/subject/grade.

In URM, as in MRM, a value-added measure of zero does not indicate that students made no growth. In URM, a value-added measure of zero indicates that students made the expected progress or similar progress as students with the average district, school, or teacher (depending on the model). URM does not use NCEs, and its value-added measure is always reported in the scaling units of the current test.

Why does EVAAS use the intra-year approach?

EVAAS recommends the intra-year approach to test transitions based on its extensive experience in test transitions over the past 20 years. To summarize the key points made previously:

- The intra-year approach does not assume that scaling is the same from one year to the next, so it is the way to use all testing data and uphold continuity of reporting as Ohio transitions to a new test.
- The intra-year approach avoids the need to "double test" students in order to equate the scales of the old and new tests because the intra-year approach is based on any change in the students' position in the distribution of NCEs from one year to the next rather than equating the scales from one year to the next.

How will EVAAS models calculate value-added measures in a transition year?

In a transition year, EVAAS will continue to use all available student test scores across all grades and subjects for the most recent five years, and the MRM (or gain-based) approach can still be used. Although the base year approach is not appropriate (since scales will change across the transition), it is possible to use the intra-year approach because the NCE distributions for each year are based on that specific year. For example, if measuring gains from 2015 to 2016, the scores from 2015 are used to create a 2015 NCE distribution and the scores from 2016 are used to create a 2016 NCE distribution. The value-added measure for 2016 is based on the change in achievement from the 2015 NCE distribution to the 2016 NCE distribution. The intra-year approach takes the scaling units of the test out of the modeling.

Because the URM currently uses only an intra-year approach and reports value-added estimates in the scaling units of the current test, a change in testing regimes does not affect the modeling or reporting. The URM reporting can continue as always during the transition year.

It is important to note that prior to any value-added analyses with new tests, EVAAS verifies that the test's scaling properties are suitable for such reporting. The new test must have sufficient stretch in the scales and must be highly related to academic standards. In addition, EVAAS verifies that the new test is related to the old test to ensure that the comparison from one year to the next is statistically reliable. Perfect correlation is not required, but there should be some relationship between the new test and old test. For example, a new grade 6 Math exam should be correlated to previous Math scores in grades 4 and 5 and to a lesser extent other grades and subjects such as Reading and Science. Once suitability of any new assessment has been confirmed, it is possible to use both the historical testing data and the new testing data to avoid any breaks or delays in value-added reporting.

Does the release of new performance levels affect value-added reporting?

No. The release of new standards or performance levels does not affect value-added reporting for students. Value-added measures are calculated from actual student scores; they make no use of the performance levels.