SAS® EVAAS®

**Misconceptions about Value-Added Reporting in Ohio**
Introduction
Since 2002, SAS® EVAAS® has provided Ohio educators and policymakers with a powerful tool to determine—grade-by-grade and subject-by-subject—whether all students have plentiful choices and increased opportunities for learning. SAS EVAAS analyses follow the progress of individual students over time to:

- Assess districts’ and schools’ influence on student progress; and
- Provide trajectories for individual students toward critical academic benchmarks.

Through the Ohio Department of Education (ODE), this reporting is available to every district, public school, and charter school in the state testing with Ohio Achievement Assessments (OAA) via a secure Web application.

The value-added estimates provided by SAS EVAAS are based on a robust and reliable methodology. This important approach overcomes many critical statistical issues related to using standardized tests to assess student progress, and mitigates concerns about fairness. This document addresses the following misconceptions about the current SAS EVAAS reporting for ODE Accountability.

**Misconceptions related to the student population served by educators**
Student growth is correlated with certain demographic variables, so EVAAS should control for demographics...

If students are already high achieving, it is harder to show growth...

EVAAS should always indicate growth if the percentage of students scoring proficient or above increased since last year...

EVAAS cannot measure the progress of districts and schools with high mobility rates...

EVAAS cannot measure growth for groups of students who have missing data...

**Misconception related to the test used in value-added modeling**
EVAAS reporting is not reliable or valid since it is based only on the OAA...

**Misconceptions related to the value-added modeling approach itself**
EVAAS is based on a “black box” methodology...

The EVAAS methodology is too complex; a more simple approach to measuring district and school effectiveness would provide better information to educators...

Growth is calculated based on how other schools perform each year...

Teacher value-added estimates are not reliable enough to be used in high-stakes decisions...
Misconceptions related to the student population served by educators

1. Student growth is correlated with certain demographic variables, so EVAAS should control for demographics.

It is widely known that students with certain socioeconomic or demographic (SES/DEM) characteristics tend to score lower, on average, than students with other SES/DEM characteristics, and there is concern that educators serving those students could be systematically disadvantaged in the modeling.

However, this adjustment is not statistically necessary for the most sophisticated value-added models, such as those used by SAS EVAAS in the State of Ohio. This is because EVAAS uses all available testing history for each individual student and does not exclude students who have missing test data. In essence, each student serves as his or her own control, and to the extent that SES/DEM influences persist over time, these influences are already represented in the student’s data.

EVAAS in Theory

As a 2004 Ed Trust study stated, specifically with regards to the SAS EVAAS modeling:

If a student’s family background, aptitude, motivation, or any other possible factor has resulted in low achievement and minimal learning growth in the past, all that is taken into account when the system calculates the teacher’s contribution to student growth in the present.¹

This approach has been confirmed through a variety of robust statistical analyses. In 2004, a SAS and Vanderbilt team published a study that closely examined SES/DEM adjustments and concluded:

SES and demographic covariates add little information beyond that contained in the covariance of test scores.²

This finding has been confirmed independently by prominent value-added experts who have replicated a variety of value-added models, including SAS EVAAS models. More specifically, a 2007 paper by RAND researchers J.R. Lockwood and Dan McCaffrey explicitly verified the SAS EVAAS models, citing them by name, when they wrote:

William Sanders, the developer of the TVAAS model, has claimed that jointly modeling 25 scores for individual students, along with other features of the approach is extremely effective at purging student heterogeneity bias from estimated teacher effects... The analytic and simulation results presented here largely support that claim.³

An economist-based perspective by UCLA researchers Pete Goldschmidt, Kilchan Choi and Kyo Yamashiro provided a similar finding in their study comparing value-added models:

First, adding in an adjustment for student SES (as measured by eligibility for free- or reduced-price lunch) adds very little once a student’s initial status is controlled... This indicates that student initial status captures many of the effects that SES is attempting to measure. In other words, by controlling for initial status, the model already captures the preceding effects that SES might have on students.⁴

EVAAS in Practice
While the statistical literature presents evidence that sophisticated value-added reporting does not need to make any adjustments for student characteristics, actual data may be the most readily apparent evidence.

The graph in Figure 1 plots the percentage of tested students who are eligible for free and reduced-price lunch at each school in Ohio against the school’s growth index (the value-added estimate divided by its standard error) for OAA Mathematics in grades five through eight in 2011. Regardless of the school’s student characteristics, there is essentially no correlation to the growth index. In other words, the dots representing each school do not trend up or down as the percentage increases; the cluster of dots is fairly even across the spectrum.

FIGURE 1: OHIO GROWTH INDEX V. PERCENT TESTED FREE/REDUCED-PRICE LUNCH BY SCHOOL

Figure 2 provides similar information for the percentage of minority students. Again, there is essentially no correlation to the growth index.

FIGURE 2: OHIO GROWTH INDEX V. PERCENT TESTED MINORITY BY SCHOOL

Unintended Policy Consequences of Adjustments
In addition to being statistically unnecessary for an approach like SAS EVAAS, adjusting for demographics can have unintended negative consequences from a policy perspective by masking inequities in student academic opportunity.
Consider the following scenario: it is well documented that novice teachers are, on average, less effective than veteran teachers; it has also been documented that schools with a higher concentration of poor and minority students tend to receive a disproportionate number of beginning teachers. In this scenario, adjustment for SES/DEM variables will over-adjust the estimates and may camouflage the fact that students in certain schools are not getting an equitable distribution of the teaching talent. By excluding such adjustments, sophisticated models are better able to highlight this disparity than models that make adjustments for SES/DEM variables.

Furthermore, any adjustments for student characteristics can send mixed messages to educators. On a philosophical level, the question educators should ask is whether they should have lower expectations for a student from a poor or minority family than one from a rich or non-minority family, even when the two students have identical test scores and academic history. By adjusting for these variables, one is directly assuming that there will be different expectations for two students with the same prior achievement pattern who come from different SES/DEM communities.

---

2. If students are already high achieving, it is harder to show growth.

Educators serving high-achieving students are often concerned that their students’ entering achievement level makes it more difficult for them to show growth. However, with EVAAS, educators are neither advantaged nor disadvantaged by the type of students that they serve. The modeling reflects the philosophy that all students deserve to make appropriate academic progress each year; as such, EVAAS provides reliable and valid measures of growth for students, regardless of their achievement level.

EVAAS in Theory

The value-added models used in Ohio are designed to estimate whether students made one year’s worth of growth. For OAA in mathematics and reading, one year’s worth of growth is about maintaining achievement levels. For OAA in science, one year’s worth of growth is about meeting expected performance based on a specific group’s prior academic performance.

Furthermore, while the OAA is designed to discriminate proficient from non-proficient, OAA is also designed to have sufficient stretch to discriminate between Limited, Basic, Proficient, Accelerated, and Advanced performance levels. Accordingly, there is sufficient stretch in the OAA testing scales to measure the growth of high-achieving students.

In fact, any test that is used in EVAAS analyses must meet three criteria, and OAA meets these criteria:

- Must be aligned to curriculum standards.
- Must be reliable and valid.
- Must demonstrate sufficient stretch at the extremes.

EVAAS in Practice

Actual data may be the most readily apparent evidence. The graph in Figure 3 plots the average entering achievement for each school in Ohio against its growth index (the value-added estimate divided by its standard error) for OAA Mathematics in grades five through eight in 2011. Regardless of the school’s achievement, there is essentially no correlation to the growth index. In other words, the dots representing each school do not trend up or down as achievement increases; the cluster of dots is fairly even across the achievement spectrum.

FIGURE 3: OHIO GROWTH INDEX V. AVERAGE ACHIEVEMENT BY SCHOOL
3. **EVAAS should always indicate growth if the percentage of students scoring proficient or above increased since last year.**

Comparing the percentage of students who score proficient (or above) over time does not account for changes in achievement within performance categories. EVAAS value-added reporting follows the progress of individual students over time, regardless of their achievement level, to ensure that all students count.

**EVAAS in Theory**
Imagine the scenario below. The mathematics achievement level of Student 1 is represented by the line with the blue diamonds, and that of Student 2 is represented by the line with the orange squares. The achievement level of Student 1 has steadily increased over time while the achievement level of Student 2 has steadily decreased over time. From seventh to eighth grade, Student 1 moved from basic to proficient performance category. From seventh to eighth grade, Student 2 maintained his position in the proficient performance category, although his achievement level has gone down.

**FIGURE 4: STUDENT TESTING HISTORY IN OAA MATHEMATICS FOR STUDENT 1 AND STUDENT 2**

Just by considering the number of students who have scored proficient, assuming all other students have maintained the same performance categories, the number of students has increased with the addition of Student 1. However, this does not take into account that Student 2’s achievement level is steadily decreasing over time. A more subtle approach is required that considers the growth of all students, regardless of their achievement level.

**EVAAS in Practice**
EVAAS does not measure students’ progress based on the number or percentage of students who tested proficient, accelerated, or advanced, as compared to previous years. EVAAS detects these subtle changes in progress, even within performance levels. As a result, educators are recognized when they make progress with students outside the “bubble.”
4. **EVAAS cannot measure the progress of districts and schools with high mobility rates.**

EVAAS value-added analyses provide reliable and valid estimates of the effectiveness of districts and schools, including those with high mobility. This is because EVAAS can include students even if they have missing test data, so that the progress of districts and schools is representative of the students served.

**EVAAS in Theory**

Highly mobile students are more likely to be low-achieving students, and it is important to include these students to avoid selection bias, which could provide misleading growth estimates to districts and schools. While more simplistic value-added or growth estimates may require that students have the same set of predictors or that students have all required predictors, this often has the result of excluding mobile student populations, and this would disproportionately affect educators serving those types of students.

EVAAS does not require that students have the same set of predictors or all required predictors, and this approach includes more students in the growth measures. When estimating students’ entering achievement, the modeling considers the quantity and quality of information available to each student, as well as student mobility among schools from year to year.

Furthermore, it is important from a philosophical perspective that as many students as possible be included in the district and school growth measures so that highly mobile student populations receive the same level of attention as non-mobile ones.

**EVAAS in Practice**

For OAA mathematics and reading, all students are included, regardless of their testing history, their number of prior test scores, and which test scores they have. For OAA Science, all students are included, so long as they have three prior test scores in any test, grade, and subject.

Because EVAAS reporting is available statewide in Ohio, students and their test history can be tracked as they move within the state.
5. EVAAS cannot measure growth for groups of students who have missing data.

EVAAS can include students even if they have missing test data, and this is a critical advantage to a sophisticated value-added approach.

**EVAAS in Theory**

Students with missing test scores are more likely to be low-achieving students, and it is important to include these students to avoid selection bias, which could provide misleading growth estimates to districts and schools that serve low-achieving or highly mobile populations of students. While more simplistic value-added or growth estimates may require that students have the same set of predictors or that students have all required predictors, this often has the result of excluding certain kind of students, and this would disproportionately affect educators serving those types of students.

EVAAS does not require that students have the same set of predictors or all required predictors, and this approach includes more students in the growth measures. When estimating students’ entering achievement, the modeling considers the quantity and quality of information available to each student, as well as student mobility among schools from year to year.

To accomplish this without imputing student test scores, EVAAS uses a sophisticated modeling approach that provides more reliable estimates of growth.6

As a simple example, consider the following scenario. Ten students are given a test in two different years. The goal is to measure academic growth (gain) from one year to the next. The right side of Figure 5 shows the same students, some of whom now have missing scores. Two simple approaches when data are missing are to calculate the mean of the differences, or to calculate the differences of the means. When there are no missing data, these two simple methods provide the same answer (5.8 on the left in the table). However, when there are missing data, each method provides a different result (9.6 vs. 4.0 on the right).

**FIGURE 5: EXAMPLE OF STUDENTS WITH MISSING DATA**

<table>
<thead>
<tr>
<th>Student</th>
<th>Fourth Grade</th>
<th>Fifth Grade</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51.9</td>
<td>74.8</td>
<td>22.9</td>
</tr>
<tr>
<td>2</td>
<td>37.9</td>
<td>46.5</td>
<td>8.6</td>
</tr>
<tr>
<td>3</td>
<td>55.9</td>
<td>61.3</td>
<td>5.4</td>
</tr>
<tr>
<td>4</td>
<td>52.7</td>
<td>47.0</td>
<td>-5.7</td>
</tr>
<tr>
<td>5</td>
<td>53.6</td>
<td>50.4</td>
<td>-3.2</td>
</tr>
<tr>
<td>6</td>
<td>23.0</td>
<td>35.9</td>
<td>12.9</td>
</tr>
<tr>
<td>7</td>
<td>78.6</td>
<td>77.8</td>
<td>-0.8</td>
</tr>
<tr>
<td>8</td>
<td>61.2</td>
<td>64.7</td>
<td>3.5</td>
</tr>
<tr>
<td>9</td>
<td>47.3</td>
<td>40.6</td>
<td>-6.7</td>
</tr>
<tr>
<td>10</td>
<td>37.8</td>
<td>58.9</td>
<td>5.8</td>
</tr>
<tr>
<td>Mean</td>
<td>50.0</td>
<td>55.8</td>
<td>5.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student</th>
<th>Fourth Grade</th>
<th>Fifth Grade</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>37.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>55.9</td>
<td>61.3</td>
<td>5.4</td>
</tr>
<tr>
<td>4</td>
<td>52.7</td>
<td>47.0</td>
<td>-5.7</td>
</tr>
<tr>
<td>5</td>
<td>53.6</td>
<td>50.4</td>
<td>-3.2</td>
</tr>
<tr>
<td>6</td>
<td>23.0</td>
<td>35.9</td>
<td>12.9</td>
</tr>
<tr>
<td>7</td>
<td>77.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>64.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>47.3</td>
<td>40.6</td>
<td>-6.7</td>
</tr>
<tr>
<td>10</td>
<td>37.8</td>
<td>58.9</td>
<td>5.8</td>
</tr>
<tr>
<td>Mean</td>
<td>45.0</td>
<td>54.6</td>
<td>4.0</td>
</tr>
</tbody>
</table>

The problem of missing data is very common to student testing data and must be taken into consideration. As illustrated above, a more sophisticated model is needed to address this problem. The approach used by EVAAS estimates the means in each of these cells using relationships between students’ test scores as if there were no missing test scores. In this way, the model provides more reliable and less biased growth measures without imputing any data. Furthermore, EVAAS uses much more student data to obtain these relationships in the growth estimates for districts and schools.

**EVAAS in Practice**

For OAA Mathematics and Reading, all students are included, regardless of their testing history, their number of prior test scores, and which test scores they have. For OAA Science, all students are included, so long as they have three prior test scores in any test, grade, and subject.

Because EVAAS reporting is available statewide in Ohio, students and their test history can be tracked as they move within the state.

Furthermore, it is important from a philosophical perspective that as many students as possible be included in the district and school growth measures so that highly-mobile student populations receive the same level of attention as non-mobile ones.
6. **EVAAS reporting is not reliable or valid since it is based only on the OAA.**

Educators may be concerned that value-added reporting relies on the use of standardized tests, which have limitations themselves. Perhaps they feel that the test does not correlate well with the curriculum or that there isn’t sufficient stretch to measure progress of very low- or high-achieving students. However, EVAAS estimates use a sophisticated modeling approach to address many of the concerns of using standardized tests, and SAS reviews the test scores annually to ensure that they are an appropriate use for EVAAS value-added reporting.

**EVAAS in Theory**

Student test scores are the basic ingredient of all SAS EVAAS analyses. SAS EVAAS is not involved in, and has no control over, test construction. OAA performs a universal assessment of Ohioan standards, and OAA exams are aligned to the appropriate grade-level standards that are sufficient for longitudinal modeling and prediction. Regardless, before using any tests in EVAAS modeling, rigorous data processing and analyses verify that the tests meet the following three criteria:

- Must be aligned to curriculum standards.
- Must be reliable and valid.
- Must demonstrate sufficient stretch at the extremes.

To date, OAA has met these criteria. More specifically, SAS analyses verify that there are enough different scaled scores at the top and bottom of the scales to differentiate student achievement. This processing also analyzes the percentage of students scoring at the top and bottom scores to ensure there are no ceilings or floors. After all analyses are completed and EVAAS estimates are available, SAS verifies that districts, schools and teachers serving both high and low achieving students can show both high and low growth. This process is repeated every year.

**EVAAS in Practice**

Actual data may be the most readily apparent evidence. The graph in Figure 6 plots the average entering achievement for each school in Ohio against its growth index (the value-added estimate divided by its standard error) for OAA Mathematics in grades five through eight in 2011. The graph demonstrates that schools serving both high- and low-achieving students can show both high and low growth, as measured by EVAAS.

**FIGURE 6: OHIO GROWTH INDEX V. AVERAGE ACHIEVEMENT BY SCHOOL**
7. EVAAS is based on a “black box” methodology.
The EVAAS methodologies and algorithms are published and have been in the open literature for 15 years. For those interested in learning more about the statistical models used in EVAAS reporting, including Ohio, the following references will be useful:


- On the **Tennessee Value-Added Assessment System**: Millman, J. (ed.) (1997). “Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?” Chapters 12 through 16 focus on the Education Value-Added Assessment System (EVAAS) upon which Ohio’s reporting is based.

**EVAAS in Theory**
While SAS EVAAS reporting benefits from a robust modeling approach, this statistical rigor is necessary to provide reliable estimates. More specifically, the SAS EVAAS models attain their reliability by addressing critical issues related to working with student testing data, such as students with missing test scores and the inherent measurement error associated with any test score.

Regardless, the SAS EVAAS modeling has been sufficiently understood such that value-added experts and researchers have replicated the models for their own analyses. In doing so, they have validated and reaffirmed the appropriateness of the SAS EVAAS modeling. The references below include recent studies by statisticians from the RAND Corporation, a non-profit research organization:


**EVAAS in Practice**
EVAAS uses multiple statistical models based on the objectives of the analyses and the characteristics and availability of the assessment data used.

- The multivariate response model (MRM) used in value-added analyses is a multivariate, longitudinal, linear mixed model. In other words, it is conceptually a multivariate repeated-measures ANOVA model. The MRM is used when scores are scaled or transformed so that the difference between two scores is meaningful. The MRM is used when there are clear “before” and “after” assessments in which to form a reliable gain estimate. In Ohio, this is used for math and reading, grades four through eight.

- The univariate response model (URM) used in value-added analyses is conceptually an analysis of covariance (ANCOVA) model. The URM is used when the test data do not meet the requirements for MRM analyses as stated above. In Ohio, this is used in subjects where the OAA is not tested in consecutive grades, such as science.
8. The EVAAS methodology is too complex; a more simple approach to measuring district and school effectiveness would provide better information to educators.

Although conceptually easy, the statistical rigor necessary to provide precise and reliable growth measures requires that several important analytical problems be addressed when analyzing longitudinal student data, which is critically important in any reporting used for educator evaluations.

In short, a simple gain calculation does not provide a reliable estimate of educator’s effectiveness. Value-added estimates based on simple calculations are often correlated with the type of students served by the educators, rather than the educator’s effectiveness with those students. Such models often unfairly disadvantage educators serving low-achieving students and unfairly advantage educators serving high-achieving students.

However, it is not necessary to be a statistician to understand the educational implications of EVAAS reporting. With the EVAAS Web application, educators have a wealth of reports that go beyond a single estimate of effectiveness and assist in identifying accelerants and impediments to student learning.

EVAAS in Theory

Any student growth or value-added model must address the following considerations in a statistically robust and reliable approach:

- **How to dampen the effects of measurement error**, which is inherent in all student assessments because the tests themselves are estimates of student knowledge, not an exact measurement.
- **How to accommodate students with missing test scores** without introducing major biases by eliminating the data for students with missing scores, using overly simplistic imputation procedures, or using very few test scores for each student.
- **How to exploit all of the longitudinal data for each student when all of the historical data are not on the same scale**.
- **How to use historical data when testing regimes have changed over time** to provide educational policymakers flexibility.

EVAAS modeling approaches address all of these concerns to provide reliable estimates of educator effectiveness, and more details are provided below.

- **EVAAS value-added measures are based on all of a student’s previous years’ performance data on an assessment instrument (rather than just one or two years of data in one or two subjects) to determine the teacher/school/district’s estimated impact on its students’ academic progress**. The inclusion of multiple years of data from multiple subjects for each individual student adds to the protection of an educational entity from misclassification in the value-added analysis. More specifically, using all available data at the individual student level can dampen the effect of measurement error, which is inherent in any test score and in all value-added or growth models.

- **EVAAS value-added measures are sophisticated and robust enough to include students with missing data**. Since low-achieving students are more likely to miss tests than high-achieving students, the exclusion of students with missing test scores can introduce selection bias, which would disproportionately affect educators serving those students.

- **EVAAS value-added measures provide estimates whether, on average, the students fell below, met, or exceed the established expectation for improvement in a particular grade/subject**. Assessing the impact at the group level, rather than on individual students, is a more statistically reliable approach, due to the issues with measurement error.
• **EVAAS value-added measures** take into account the measures of uncertainty (standard error) when determining whether an educational entity is decidedly above or below expected progress, as defined by the model. Any model based on assessment data relies on estimates of student learning, and it is important that any value-added measure take into account the inherent uncertainty when providing estimates.

• **EVAAS value-added models are sophisticated enough to accommodate different tests or changes in testing regimes.** This provides educators with additional flexibility. First, they can use more tests, even if they are on differing scales. Second, they can continue to provide reporting when the tests change, as will be the case with the implementation of the Common Core State Standards and the PARCC assessments.

SAS EVAAS statistical models have been validated and vetted by a variety of value-added experts. The references below include recent studies by statisticians from the RAND Corporation, a non-profit research organization:


**EVAAS in Practice**

Although the statistical approach is robust and complex, the reports in the EVAAS Web application are easy to understand. Provided by subject, grade, and year, the value-added estimates are color-coded like a traffic light: green indicates that students in a district or school made more than the expected progress; yellow indicates that students in a district or school made about the expected progress; and red indicates that students in a district or school made less than the expected progress. Educators and administrators can identify their strengths and opportunities for improvement at a glance. The reporting is interactive, so that authorized users can drill down to access diagnostic reports for students by subgroup or achievement level, individual student-level projections, and other reports. Educators have a comprehensive view of past practices as well as tools for current and future students. Thus, educators benefit from the rigor of the SAS EVAAS models by gaining insight in an accessible and non-technical format.
FIGURE 7: SAMPLE SAS EVAAS DISTRICT VALUE-ADDED REPORT

<table>
<thead>
<tr>
<th>Report: District Value-Added</th>
<th>Test: OAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>District: District 1</td>
<td>Subject: Mathematics</td>
</tr>
<tr>
<td>Year: 2011</td>
<td>Type: Accountable</td>
</tr>
</tbody>
</table>

Please note the OAA Tested reports include students that were tested at the school or district and not only those students that were accountable to that school or district. Reporting including only those students that were accountable is available under the OAA Accountable test.

The values in the table below are rounded for display purposes.

<table>
<thead>
<tr>
<th>Estimated District Mean NCE Gain</th>
<th>Mean NCE Gain over Grades Relative to</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Growth standard</td>
</tr>
<tr>
<td>Grade</td>
<td>3</td>
</tr>
<tr>
<td>Growth Standard</td>
<td>0.0</td>
</tr>
<tr>
<td>State 3-Yr-Avg</td>
<td></td>
</tr>
<tr>
<td>2009 Mean NCE Gain</td>
<td></td>
</tr>
<tr>
<td>Std Error</td>
<td></td>
</tr>
<tr>
<td>2010 Mean NCE Gain</td>
<td>-1.0R</td>
</tr>
<tr>
<td>Std Error</td>
<td>0.3</td>
</tr>
<tr>
<td>2011 Mean NCE Gain</td>
<td>-3.6R</td>
</tr>
<tr>
<td>Std Error</td>
<td>0.3</td>
</tr>
<tr>
<td>3-Yr-Avg NCE Gain</td>
<td></td>
</tr>
<tr>
<td>Std Error</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimated District Mean NCE Scores</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade</td>
<td>3</td>
</tr>
<tr>
<td>State Base Year (2010)</td>
<td>60.0</td>
</tr>
<tr>
<td>State 3-Yr-Avg</td>
<td></td>
</tr>
<tr>
<td>2008 Mean</td>
<td></td>
</tr>
<tr>
<td>2009 Mean</td>
<td>37.1</td>
</tr>
<tr>
<td>2010 Mean</td>
<td>37.5</td>
</tr>
<tr>
<td>2011 Mean</td>
<td>35.5</td>
</tr>
</tbody>
</table>

| G | Estimated mean NCE gain is above the growth standard by at least 2 standard errors. |
| Y | Estimated mean NCE gain is below the growth standard by at least 2 standard errors but less than 2 standard error above it. |
| R | Estimated mean NCE gain is below the growth standard by more than 2 standard errors. |

To view additional reports, click on the underlined numbers or words.
**9. Growth is calculated based on how other schools perform each year.**

Growth (grades four through eight in math and reading) is not based on the performance of other schools each year. Rather, growth of a group of students is based on the group’s change in achievement from one school year to the next based on the 2010 OAA achievement distributions. The 2010 OAA Math and Reading distributions provide typical demonstrations of achievement of groups of students as they progress through the grade levels. EVAAS evaluates the score distributions on an annual basis to verify that using the base year of 2010 continues to be appropriate.

**EVAAS in Theory**

Performance of other groups in a given year does not affect the growth calculation of the cohort in question. Each group becomes its own control group.

Students’ changes in achievement level over time can be compared to students’ changes in height over time:

- For a student to get taller, another student does not have to get shorter.
- A student can grow taller in a given year, no matter how his/her peers grow.

**EVAAS in Practice**

In grades four through eight for reading and mathematics, a group of students makes one year’s worth of academic growth in a school year when the group maintains the same relative position in the distribution in the current year as compared to the previous year, using the 2010 base year distributions.

Figures 8 and 9 below provide the OAA scale score distributions for fifth and sixth grade reading in the base year. The arrow indicates how a school might have changed its position in the distribution from one year to the next.

**FIGURE 8: FIFTH GRADE READING 2010 DISTRIBUTION**

**FIGURE 9: SIXTH GRADE READING 2010 DISTRIBUTION**

Each year, a group’s growth is calculated by comparing its position in the current grade distribution from 2010 to its former position in the previous grade’s distribution from 2010. In this example, comparing the placement of the arrow in each graph shows that the school has moved up in the distribution, as compared to the base year.
Misconceptions related to the value-added modeling approach itself

10. Teacher value-added estimates are not reliable enough to be used in high-stakes decisions.

Many studies on teacher estimates focus on single-year estimates, some of which are derived from simplistic value-added or growth models. However, SAS EVAAS teacher value-added estimates are based on a very robust statistical approach and report a multiple-year average, whenever available. The approach provides very reliable teacher estimates, which educators can use for a variety of educational and policy decisions.

EVAAS in Theory

Many critics use the repeatability of teacher value-added estimates as a proxy for their reliability. However, “perfect” repeatability is not the goal, as some year-to-year variation among individual teachers’ estimates is to be expected. Cohorts of students change every year and teachers may be more effective with one group than other. Also, some teachers may improve, or worsen, in their effectiveness over time. However, the presence of strong reliability indicates that teachers’ value-added estimates are related to their consistent skills and are not generated primarily from a random component.

SAS research and analysis has compared a variety of modeling approaches and found the following:

- The EVAAS methodology used in Ohio yields repeatability estimates around 0.70 and 0.80 for three-year teacher estimates. The repeatability estimate and residual coefficients (a measure of bias) taken together suggest that the estimates themselves are mostly related to the actual teacher’s effectiveness.\(^7\)

- Not all value-added approaches are created equal. More simplistic value-added approaches do not yield robust repeatability estimates and residual coefficients.

This has enormous implications in terms of the usefulness of the reporting provided by SAS EVAAS: educators and policymakers can rely on the teacher estimates to inform their decisions.

EVAAS in Practice

This reliability does not simply exist in a research setting. SAS EVAAS teacher value-added estimates, derived from the same approach as is provided in Ohio and provided to teachers in another state over the span of 14 years, have similar repeatability. An analysis of their estimates yielded important insights into how different kinds of teachers may change in effectiveness over time. More specifically:

- Highly effective teachers are very likely to remain effective. Teachers identified as highly effective after their first three years of teaching were extremely likely to remain highly effective three years into the future (about 95% were either average or above average in effectiveness).

- Less effective teachers may improve over time. For the teachers identified as ineffective based on three-year estimates, approximately half of them will continue to be identified as ineffective three years later.

Thus, if policymakers, administrators, and educators make high-stakes decisions based on three-year estimates, there is very little risk that the teachers identified as effective will be identified as ineffective three years later. However, additional flexibility or time may be required to assess which teachers will improve in contrast with those who do not.

In other words, in using a robust and reliable statistical approach, like SAS EVAAS, for teacher estimates, Ohioan educators and policymakers can build insightful policies customized to the teachers in their schools, districts, and state.

\(^7\)Publication forthcoming, 2012