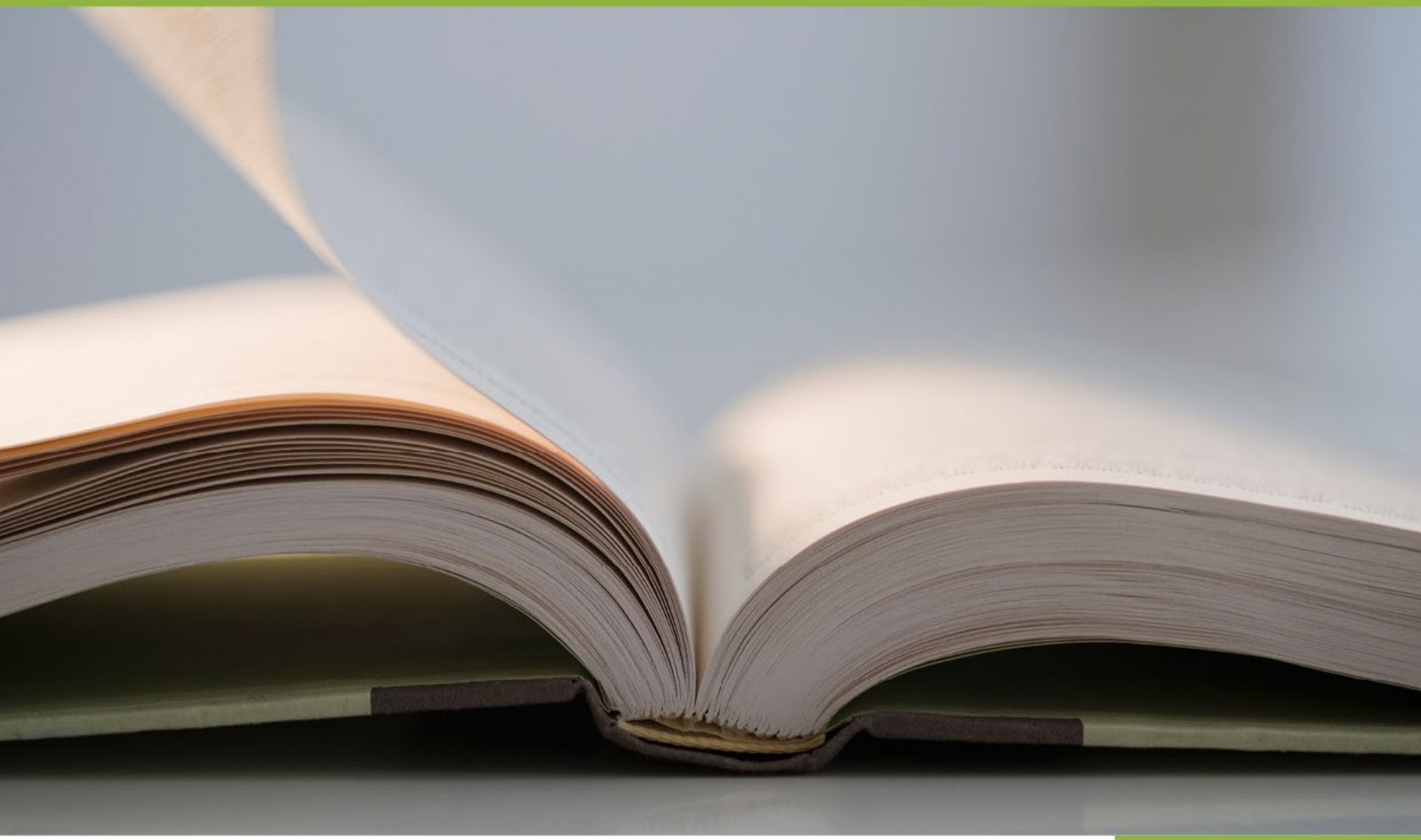
SAS® EVAAS

Topics in Value-Added Modeling in Tennessee



sas
THE POWER TO KNOW.

Introduction

Since 1993, TVAAS has provided Tennessee 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. TVAAS analyses follow the growth of individual students over time to:

- Assess districts', schools', and teachers' influence on student growth; and
- Provide trajectories for individual students toward critical academic benchmarks.

Through the Tennessee Department of Education (TDOE), this reporting is available to every district, public school, and charter school in the state via a secure web application.

The value-added estimates provided by TVAAS are based on a robust and reliable methodology. This approach overcomes many critical statistical issues related to using standardized tests to assess student growth and mitigates concerns about fairness. The purpose of this document is to explain how the TVAAS models address common topics about measuring growth using both theoretical and empirical data.

Topics Related to the Student Population Served by Educators

Misconception: Student growth is correlated with certain demographic variables, so TVAAS 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. There is concern that educators serving those students could be systematically disadvantaged in the modeling, leading some to believe that TVAAS should control for demographics.

However, this adjustment is not statistically necessary for TVAAS because, like other sophisticated value-added models, it uses all available testing history for each individual student and does not exclude students who have missing test data. Ultimately, each student serves as their own control for SES/DEM influences, and, to the extent that SES/DEM influences persist over time, they are already represented in the student's data.

TVAAS in Theory

As a 2004 Ed Trust study stated, specifically with regard to the TVAAS modeling:

> [I]f 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:

[1] Kevin Carey, "The Real Value of Teachers: Using New Information about Teacher Effectiveness to Close the Achievement Gap," *Thinking K-16* 8, no. 1 (Winter 2004): 27.

> 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 TVAAS models. For example, a 2007 paper by RAND researchers J.R. Lockwood and Dan McCaffrey explicitly verified the TVAAS models and described them as "extremely effective" at reducing bias in estimates of teacher's contributions to growth.[3] UCLA researchers Kilchan Choi, Pete Goldschmidt, 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]

TVAAS in Practice

Although the statistical literature presents evidence that sophisticated value-added reporting does not need to make any adjustments for student characteristics, actual data offers more meaningful evidence.

The graph in Figure 1 plots the percentage of tested students who are considered economically disadvantaged at each school in Tennessee against the school's growth index for TCAP Mathematics in grades 4–8 in 2024. Regardless of the school's student characteristics, there is little to 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: Tennessee Growth Index Versus Percentage Tested Economically Disadvantaged by School

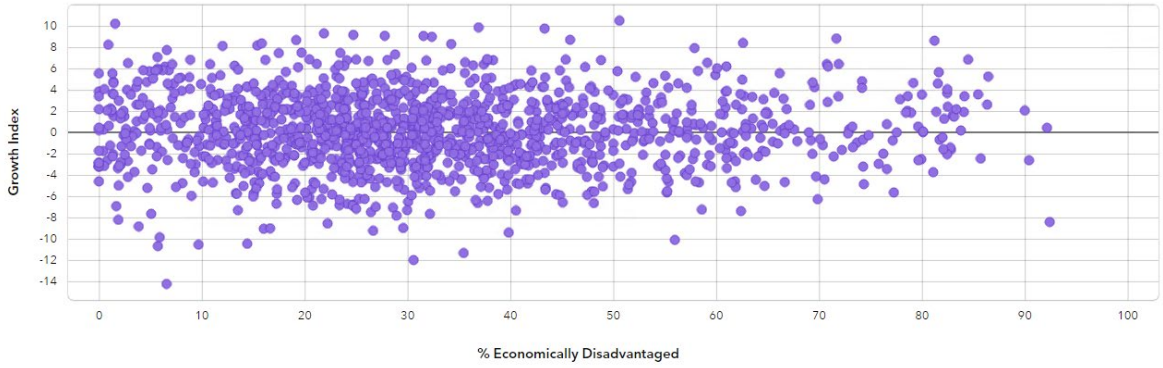


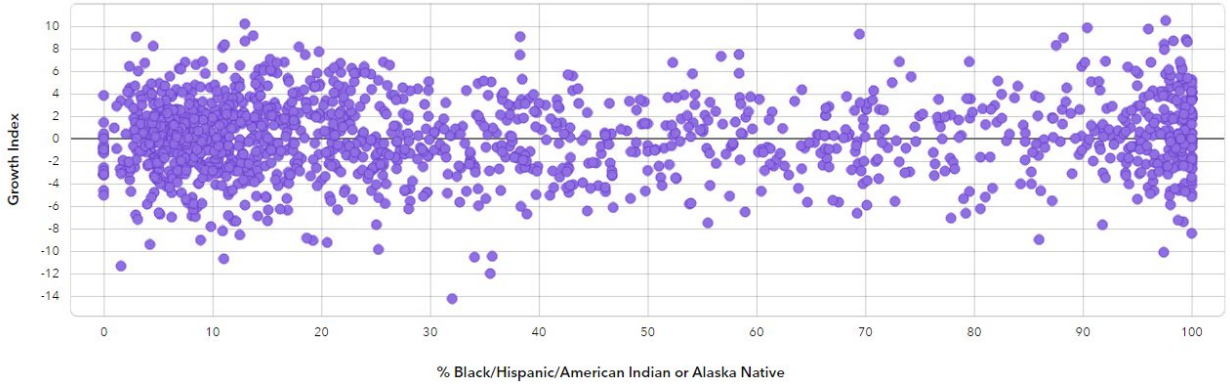
Figure 2 provides similar information for the percentage of minority students. Again, there is little to no correlation to the growth index.

[2] Dale Ballou, William Sanders, and Paul Wright, "Controlling for Student Background in Value-Added Assessment," *Journal of Education and Behavioral Statistics*, 29, no. 1 (2004): 37-65.

[3] J.R. Lockwood and Daniel F. McCaffrey, "Controlling for Individual Heterogeneity in Longitudinal Models, with Applications to Student Achievement," *Electronic Journal of Statistics*, 1 (2007): 244.

[4] Kilchan Choi, Pete Goldschmidt, and Kyo Yamashiro, *Exploring Models of School Performance: From Theory to Practice (CSE Report 673)* (Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), 2006), 24.

Figure 2: Tennessee Growth Index Versus Percentage Tested Minority by School



Misconception: If students are already high (or low) achieving, it is harder to show growth.

Educators serving either high- or low-achieving students are often concerned that their students' entering achievement level makes it more difficult for them to show growth. However, with TVAAS, 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 growth each year; as such, TVAAS provides reliable and valid measures of growth for students regardless of their achievement level.

TVAAS in Theory

The value-added models used in Tennessee are designed to follow the progress of individual students over time and estimate whether these students made the average amount of growth observed in the state in the current year for the subject (for EOC) or subject/grade (for TCAP) of interest. Although TCAP and the EOCs are designed to discriminate proficiency from non-proficiency, they are also designed to have sufficient stretch to measure student performance at a wide range of achievement levels. Accordingly, there is sufficient stretch in the TCAP and EOC assessment testing scales to differentiate performance and measure the growth of both high- and low-achieving students.

It is a requirement that any test that is used in TVAAS analyses meet the criteria of demonstrating sufficient stretch at the extremes. This requirement ensures that progress can be measured for both low-achieving students as well as high-achieving students.

Some educators are concerned about their students who make perfect scores and how that might impact their value-added reporting. This concern primarily comes into play among students who make perfect scores year after year and are therefore unable to demonstrate growth. In truth, very few students make perfect scores in the same subject from year to year. In 2024, the number of students who made a perfect score in consecutive years for TCAP Mathematics was a fraction of a percentage—only 0.119%, or approximately 1 in 1,000. In the other TCAP subjects it was even less, with values of 0.008% in ELA, 0.007% in Science, and 0.011% in Social Studies.

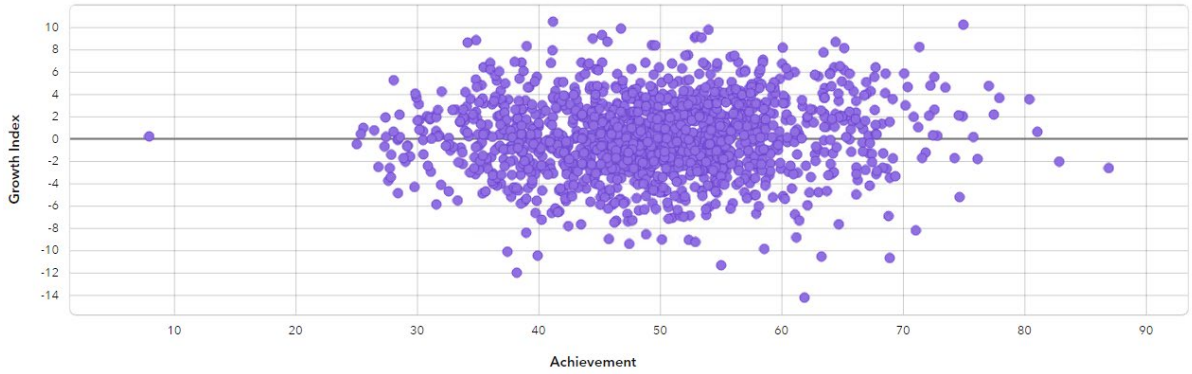
Similarly, some educators are concerned about their students who make very low scores and how that might impact their value-added reporting. TVAAS is focused on *growth* rather than *achievement*, and this approach uses multiple years of data, when available, to follow the progress of individual students over time. The growth model itself assesses whether, on average, the achievement for a group of

students increased, decreased, or stayed about the same over a period of time. This can happen regardless of whether students' prior achievement was relatively low, middle, or high. In other words, educators are not disadvantaged by serving low-achieving students who are not yet proficient.

TVAAS in Practice

Actual data shows that schools' value-added measures are not typically related to their students' achievement level. The graph in Figure 3 plots the average entering achievement for each school in Tennessee against its growth index for TCAP Mathematics in grades 4–8 in 2024. Regardless of the school's achievement, there is little to 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: Tennessee Growth Index Versus Average Achievement by School



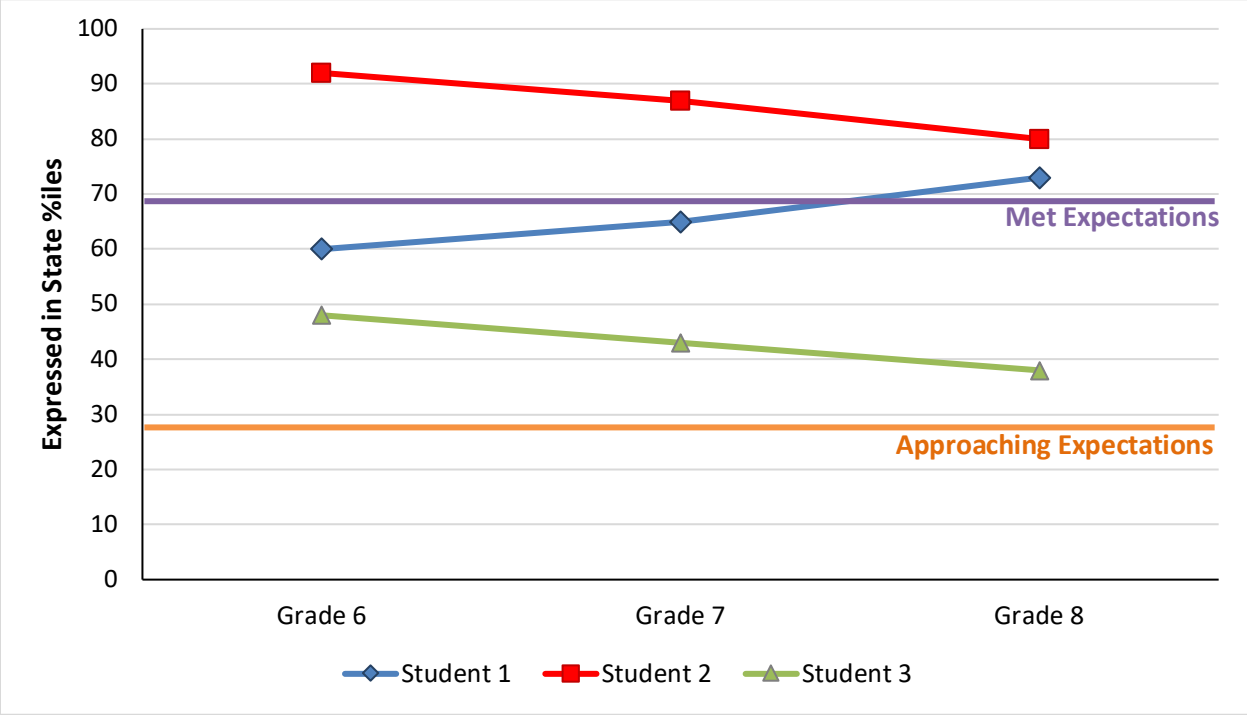
Misconception: TVAAS should always indicate growth if the percentage of students scoring Met Expectations or above increased since last year.

Academic proficiency is in an important metric for measuring student success, and increasing the number of proficient students is a goal of many educational entities. However, measuring only the percentage of students who score Met Expectations (or above) over time does not account for changes in achievement within performance categories. Additionally, comparing the proficiency rate at a district or school over time does not account for changes in the cohort of students included in the rate. TVAAS value-added reporting follows the progress of individual students over time, regardless of their achievement level, to ensure that all students count.

TVAAS in Theory

Consider the achievement history of three students illustrated in Figure 4. The achievement level of Student 1 has steadily increased over time, and between seventh and eighth grades, Student 1 moved from the Approaching to Met Expectations performance category. Alternatively, the achievement of Student 2 has been steadily declining. Despite this drop in absolute achievement, Student 2 maintained their position in the Met Expectations performance category. Student 3 has had steadily declining performance and has stayed below the Met Expectations threshold.

Figure 4: Student Testing History in ELA for Student 1, Student 2, and Student 3



By considering only the change in proportion of students who have Met Expectations, it would appear that this group of three students is generally improving because the number of students meeting expectations has increased with the addition of Student 1. However, this does not consider that the achievement level of the other two students is steadily decreasing over time. Alternatively, looking at growth would indicate that, as a group, these students are likely not making the expected growth. This example helps illustrate the additional insights gained from a growth model that considers the changes in achievement for all students.

TVAAS in Practice

TVAAS does not measure students' growth based on the number or percentage of students who tested Met Expectations or Exceeded Expectations as compared to previous years. TVAAS detects changes in growth both across and within performance levels. As a result, TVAAS captures growth made by all students, even those outside the "bubble" near the proficiency threshold.

Along these lines, it is important to remember that the proficiency rate and growth rate are two different metrics; each is designed to show a different perspective on students' academic experiences. Proficiency rates provide a snapshot of students' knowledge at a specific point in time whereas TVAAS provides a measure of students' change in achievement over time. The district/school proficiency rate might increase or decrease from one year to the next due to changes in the student population associated with the district/school rather than an increase or decrease in students' growth. For example, some students might transfer in and out of the district/school. In fact, an entire grade of students will enter the district/school calculation (such as third graders) while an entire grade of students will leave the district/school calculation (such as eighth graders). As a result, changes in proficiency over time might not align to changes in growth over time.

Misconception: TVAAS cannot measure growth for groups of students who have missing data.

Some measures of student growth cannot account for students with missing data. However, TVAAS can include students even if they have missing test data, and this is a critical advantage to a sophisticated value-added approach.

TVAAS in Theory

Including students with missing test scores as part of growth estimates is important for accurate and meaningful growth estimates because, in practice, students with missing test scores are more likely to be low-achieving students. Excluding these students from analysis could provide misleading growth estimates to systems and schools that serve low-achieving or highly mobile populations of students. Although more simplistic value-added or growth estimates might require that students have the same set of prior test scores or that students have all required prior test scores, this often has the result of excluding students with certain characteristics, and this would disproportionately affect educators serving those students.

To counteract this, TVAAS does not require that students have the same set of prior test scores or all required prior test scores, thus including 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, TVAAS uses a sophisticated modeling approach that provides more reliable estimates of growth.[5]

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 side of Figure 5). However, when there are missing data, each method provides a different result (4.6 versus 6.9 on the right side of Figure 5).

[5] S. Paul Wright, "Advantages of a Multivariate Longitudinal Approach to Educational Value-Added Assessment Without Imputation." Paper presented at National Evaluation Institute, July 8-10, 2004, Colorado Springs, CO.

Figure 5: Scores without Missing Data, and Scores with Missing Data

Student	Previous Score	Current Score	Gain
1	51.9	74.8	22.9
2	37.9	46.5	8.6
3	55.9	61.3	5.4
4	52.7	47.0	-5.7
5	53.6	50.4	-3.2
6	23.0	35.9	12.9
7	78.6	77.8	-0.8
8	61.2	64.7	3.5
9	47.3	40.6	-6.7
10	37.8	58.9	21.1
Column Mean	**50.0**	**55.8**	**5.80**
Difference between Current and Previous Score Means			**5.80**

Student	Previous Score	Current Score	Gain
1	51.9	74.8	22.9
2		46.5	
3	55.9	61.3	5.4
4		47.0	
5	53.6	50.4	-3.2
6	23.0	35.9	12.9
7	78.6	77.8	-0.8
8	61.2	64.7	3.5
9	47.3	40.6	-6.7
10	37.8	58.9	21.1
Column Mean	**51.2**	**55.8**	**6.9**
Difference between Current and Previous Score Means			**4.6**

The problem of missing data is common to student testing data and must be taken into consideration. As illustrated above, simple approaches can be problematic, and a more sophisticated model is needed to address this problem. The approach used by TVAAS 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.

TVAAS in Practice

For TCAP Mathematics and English Language Arts, all students who meet the business rules defined by TDOE are included regardless of their testing history, their number of prior test scores, and which test scores they have. For EOC and TCAP Science and Social Studies, all students are included as long as they have three prior test scores in any test, grade, and subject.

Because TVAAS reporting is available statewide in Tennessee, students and their testing 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 system and school growth measures so that highly mobile student populations receive the same level of attention as non-mobile ones.

Misconceptions Related to the Tests Used in Value-Added Modeling

Misconception: TVAAS reporting is not reliable or valid since it is based only on standardized assessments.

Educators might 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 is not sufficient stretch to measure growth of very low- or high-achieving students. However, TVAAS estimates use a sophisticated modeling approach to address many of the concerns of using standardized tests, and TVAAS reviews the test scores annually to ensure that they are an appropriate use for TVAAS value-added reporting.

TVAAS in Theory

Student test scores are the basic ingredient of all TVAAS analyses. TCAP and EOC assessments are aligned to the appropriate grade- and subject-level state standards that are sufficient for longitudinal modeling and prediction. Regardless, before using any tests in TVAAS modeling, rigorous data processing and analyses verify that the tests meet the following three criteria. The tests:

- Must be designed to assess the academic standards.
- Must be reliable and valid (usually related to the number of test questions).
- Must demonstrate sufficient stretch at the extremes.

To date, TCAP and EOC assessments have met these criteria. More specifically, TVAAS 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 TVAAS estimates are available, TVAAS 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. More details about how the models underlying TVAAS support measuring growth among high and low achieving students can be found on page 3.

Additionally, the modeling approaches that underlie TVAAS reporting are designed to minimize the impacts of the inherent measurement error associated with standardized tests on estimates of growth. First, TVAAS models use all available testing data for each student. This minimizes the impact of a single unreliable test score. Additionally, TVAAS reporting includes both estimates of growth as well as the *standard error.* The standard error is a measure of the quantity and quality of student level data included in the estimate, such as the number of students and the occurrence of missing data for those students. Taken together, the estimate and standard error provide critical information about the confidence that students are making more or less than expected growth and reduce the risk of misclassification (for example, identifying a teacher as ineffective when they are truly effective).

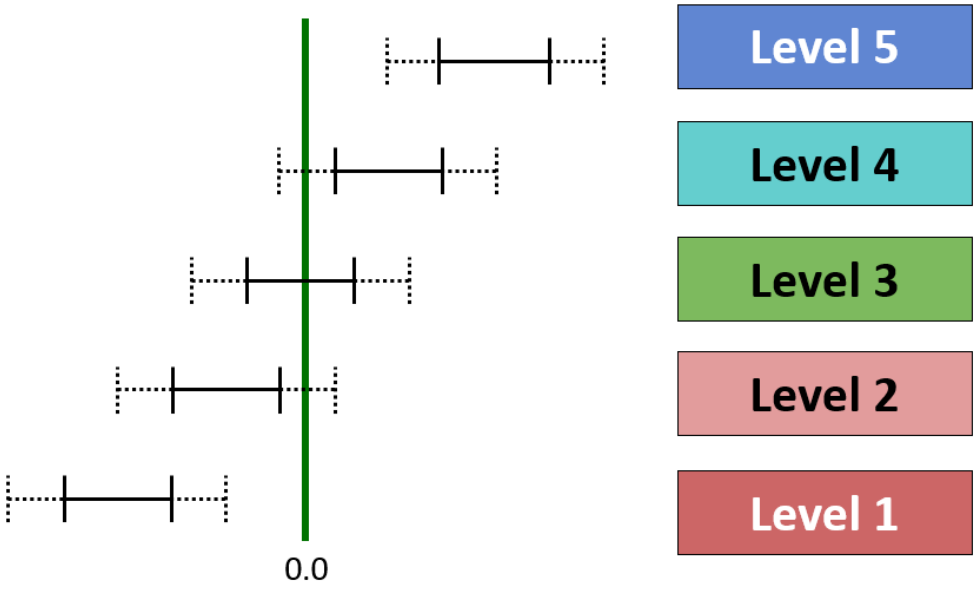
TVAAS in Practice

Each value-added estimate has an associated standard error, which is a measure of uncertainty that depends on the quantity and quality of student data associated with that value-added estimate. The standard error can help indicate whether a value-added estimate is significantly different from the growth standard. For TVAAS growth reporting, this is essentially when the growth measure is more than or less than one or two standard errors above or below expected growth or, in other words, when the

growth index is more than +1 or +2 or less than -1 or -2. These definitions then map to the growth indicators in the reports.

Figure 6 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 Level 5 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 Level 1 category. Levels 4 and 2 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. Level 3 indicates that there is evidence that students made growth as expected as both the solid and dotted cross the line indicating expected growth.

Figure 6: Visualization of Growth Categories with Expected Growth, Growth Measures, and Standard Errors



Misconceptions Related to the Value-Added Modeling Approach Itself

Misconception: TVAAS is based on a "black box" methodology.

TVAAS is based on established statistical models that have been in use among many industries for decades and, in some instances, centuries. These models are designed to work well with large amounts of information and accommodate common issues with student testing, such as non-random missing data. Although the underlying program code for these models and algorithms used for Tennessee is proprietary, the TVAAS methodologies and algorithms are published and have been in the open literature for over 25 years. Details about the TVAAS models are available in the references below:

- On the **statistical models upon which Tennessee's reporting is based**: "Statistical Models and Business Rules" available at https://tvaas.sas.com/support/TVAAS-Statistical-Models-and-Business-Rules.pdf.
- On the **Tennessee Value-Added Assessment System**: Millman, Jason (Ed.) Chapters 12-16 in *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* (Thousand Oaks, CA: Corwin, 1997).

TVAAS in Theory

Although the modeling approach underlying TVAAS can be widely understood at a high-level, a background in statistics and value-added modeling can aid in fully understanding the technical details. This statistical complexity and rigor are necessary to provide reliable growth estimates. More specifically, the TVAAS models attain their reliability by using advanced techniques to address 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.

These details have been published and made available to a community of experts for review and critique. Through this process, the TVAAS modeling has been sufficiently described 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 TVAAS modeling. The references below include studies by statisticians from the RAND Corporation, a non-profit research organization:

- On the **advantages of the longitudinal, mixed model approach**: Lockwood, J.R. and Daniel F. McCaffrey. 2007. "Controlling for Individual Heterogeneity in Longitudinal Models, with Applications to Student Achievement." *Electronic Journal of Statistics* 1: 223-52.
- On the **insufficiency of simple value-added models**: McCaffrey, Daniel F., B. Han, and J.R. Lockwood. 2008. "From Data to Bonuses: A Case Study of the Issues Related to Awarding Teachers Pay on the Basis of the Students' Progress." Presented at *Performance Incentives*: *Their Growing Impact on American K-12 Education*, Feb. 28-29, 2008, National Center on Performance Incentives at Vanderbilt University.

TVAAS in Practice

TVAAS includes two main statistical models, each described briefly below.

- The gain model (also known as the multivariate response model or MRM) used in value-added analyses is a multivariate, longitudinal, linear mixed model. The gain model is typically used when there are clear "before" and "after" assessments in which to form a reliable gain estimate. This is used for the TCAP reporting in Mathematics and English Language Arts for grades 4–8.

- The predictive model (also known as the univariate response model or URM) used in value-added analyses is conceptually an analysis of covariance (ANCOVA) model. The predictive model is based on the difference between expected and actual scores for students. In Tennessee, this is used for TCAP ELA and Math for grade 3, TCAP Science for grades 5–8, TCAP Social Studies for grades 6–8, and EOC assessments.

For more details about these modules, please visit the TVAAS e-Learning modules which can be found at: https://tvaas.sas.com/learningModules.html.

Misconception: The TVAAS methodology is too complex; a simpler approach to measuring system and school effectiveness would provide better information to educators.

There is some concern among educators and administrators that the underlying methodology for arriving at growth estimates cannot be readily understood by practitioners and that this limits the ability of those practitioners to make meaningful insights and decisions based on the reporting. However, complex modeling techniques are required to provide precise and reliable growth measures in the complex reality of statewide educational systems.

For example, 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.

Furthermore, it is not necessary to have an in-depth understanding of the modeling underlying TVAAS to use it to make effective decisions. Just as most people might not understand how meteorologists arrive at their weather forecasts but can use it to inform their decisions, educators can use the data provided by TVAAS reporting to guide their practice. With the TVAAS 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. These reports have been designed to approachable and useful for educators with or without advanced statistical backgrounds.

TVAAS 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 use 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 programs have changed over time** to provide educational policymakers flexibility.

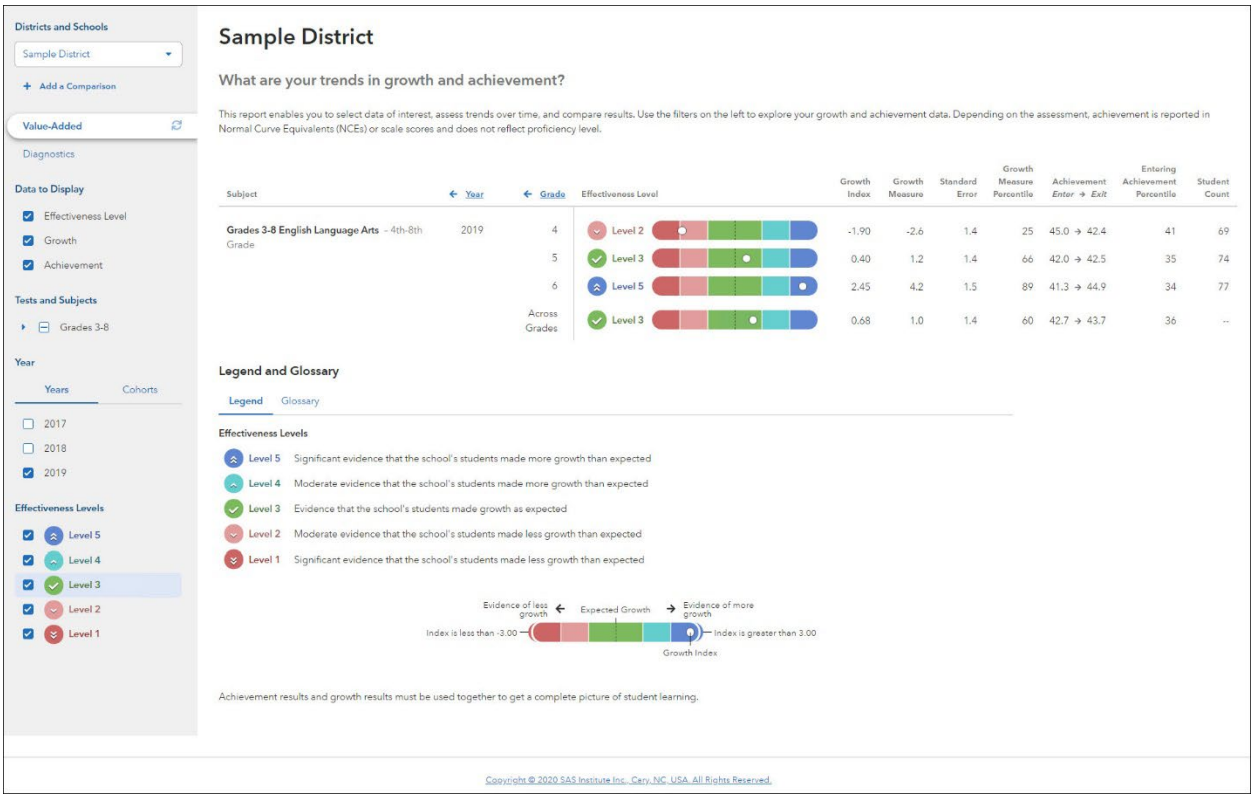
TVAAS modeling approaches address all of these concerns to provide reliable estimates of educator effectiveness. In particular:

- **TVAAS 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/system's estimated impact on its students' academic growth.** 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.
- **TVAAS 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.
- **TVAAS value-added measures provide estimates whether, on average, the students fell below, met, or exceeded 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.
- **TVAAS value-added measures consider the measures of uncertainty (standard error) when determining whether an educational entity is decidedly above or below expected growth 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 consider the inherent uncertainty when providing estimates.
- **TVAAS value-added models are sophisticated enough to accommodate different tests or changes in testing programs.** 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.

TVAAS in Practice

Although the statistical approach is robust and complex to safeguard against the issues previously described, the reports in the TVAAS web application are easy to understand. Provided by subject, grade, and year, the value-added estimates are color-coded for quick interpretation: blue indicates that students in a district or school made more than the expected growth; green indicates that students in a system or school made about the expected growth; and red indicates that students in a system or school made less than the expected growth. Educators and administrators can identify their strengths and opportunities for improvement at a glance. The reporting is wide-ranging, so that authorized users can also access Diagnostic reports for students by group 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 TVAAS models by gaining insight in an accessible and non-technical format. For more details about the reporting, please visit the TVAAS e-Learning modules found at: https://tvaas.sas.com/learningModules.html.

Figure 7: Sample TVAAS System Value-Added Report



Misconception: There is a fixed number of districts, schools, and teachers for each effectiveness level.

The TVAAS value-added measures are based on the statewide pool of test-takers for the subject/grade/year (or subject/year for EOC) of interest. More specifically, expected growth is based on a comparison against the average amount of growth observed in the state. However, there is *not* a set distribution or fixed number of districts, schools, and teachers for each effectiveness level.

TVAAS in Theory

TVAAS value-added measures are based on a comparison against the average observed growth in the state for the particular year of reporting. By comparing growth against the actual statewide performance, these models can compensate for changes that affect the entire student population, such as changes in standards, assessments, or education policy. Because growth measures are relative to the growth across the state, they tend to be centered around zero. In this case, a measure of zero represents the expected growth based on what was observed within the state. Consequently, approximately half of the district/school/teacher estimates fall above zero and approximately half of the district/school/teacher estimates fall below zero.

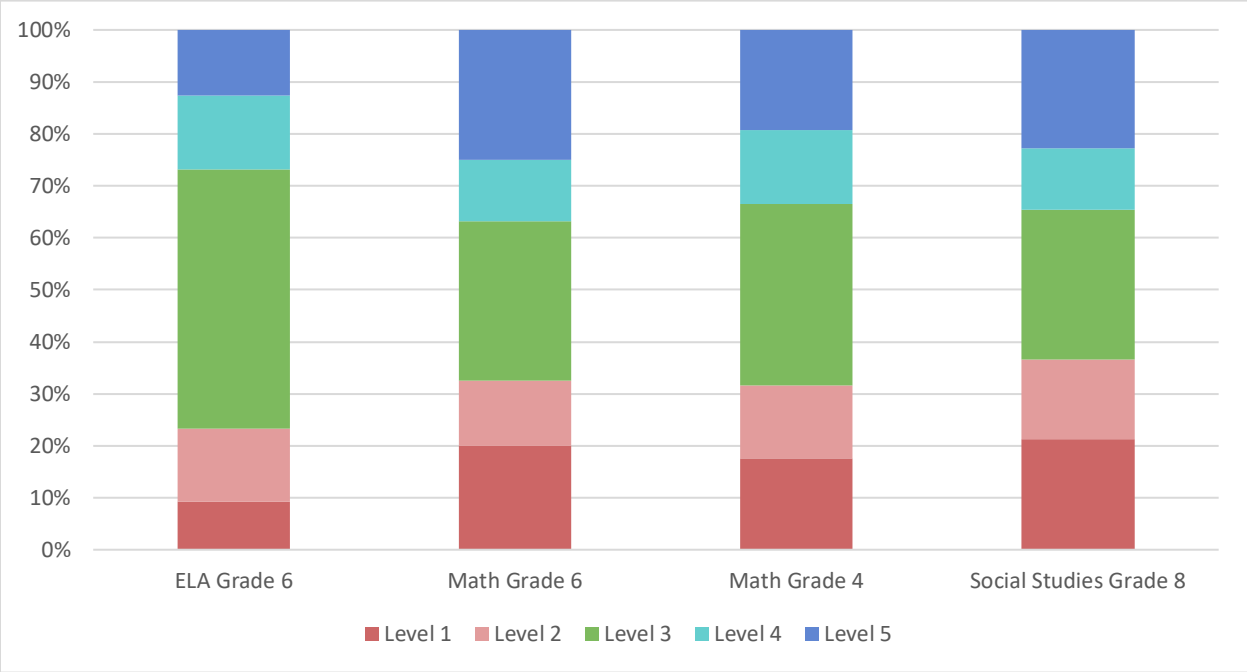
However, it should be noted that there is not a set distribution of the value-added measures and being centered on expected growth does not mean half of the measures would be in the positive levels and half would be in the negative levels since many value-added measures are indistinguishable from the expectation when considering the statistical certainly around that measure.

TVAAS in Practice

The distribution of effectiveness levels based on single-year 2024 TVAAS measures for teachers across subjects and grades is provided below, and this information illustrates that the distribution can vary by subject and grade with the majority of teachers' students meeting or exceeding expected growth. The actual distribution depends on the extent to which there are measurable differences among districts, schools, and teachers, and the exact distribution in a subject/grade can vary from year to year.

As Figure 8 below also illustrates, the percentage of reports considered either Level 1 or 2 is similar to those considered Level 4 or 5. Additionally, though these percentages are similar to one another in a given year, they do not necessarily hold constant from year to year.

Figure 8: Distribution of Teacher Effectiveness Levels by Subject and Grade



Misconception: 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, TVAAS teacher value-added estimates are based on a robust statistical approach and report a multiple-year average whenever available. The approach provides reliable teacher estimates, which educators can use for a variety of educational and policy decisions.

TVAAS 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 might be more effective with one group than another. Also, some teachers might increase or decrease 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.

TVAAS in Practice

To explore the reliability of teachers' value-added estimates, TVAAS results from the past two decades found that:

- **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 effective three years into the future (about 95% were either average or above average in effectiveness).
- **Less effective teachers might 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 while the other half will improve their effectiveness during the same period.

This has enormous implications in terms of the usefulness of the reporting provided by TVAAS. It means that educators and policymakers can rely on the teacher estimates to inform their decisions.

Additionally, in 2012, the Tennessee Department of Education reported to the legislature that less than 1% of Tennessee teachers moved from Level 5 (most effective) to Level 1 (least effective) from one year to the next based on three-year TVAAS teacher estimates. With these robust, three-year data sets, teachers and leaders have valid and reliable information on students' growth.

In other words, in using a robust and reliable statistical approach for teacher estimates, such as TVAAS, Tennessee educators and policymakers can build insightful policies customized to the teachers in their schools, systems, and state.