EVAAS

SAS® EVAAS®

*Mis*conceptions about Value-Added Reporting in Tennessee

Version 2.0 August 2015

sas Solutions OnDemand

1 Introduction

Since 1993, SAS® EVAAS® 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 progress of individual students over time to:

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

Through the Tennessee Department of Education (TDOE), this reporting is available to every system, public school, and charter school in the state 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.

*Mis*conceptions related to the student population served by educators

*Mis*conception related to the test used in value-added modeling

*Mis*conceptions related to the value-added modeling approach itself

1. 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, 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 for TVAAS in the State of Tennessee. This is because TVAAS 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.

TVAAS in Theory

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

> *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]

---

[1] Carey, K. (Winter 2004). The Real Value of Teachers: If Teachers Matter, Why Don't We Act Like It? (The Education Trust: Washington DC).

[2] Ballou, D., W. Sanders, and P. Wright. 2004. "Controlling for Student Background in Value-Added Assessment." *Journal of Education and Behavioral Statistics*, 29(1), pp. 37-65.

[3] Lockwood J.R. and D.F. McCaffrey (2007). "Controlling for individual heterogeneity in longitudinal models, with applications to student achievement." *Electronic Journal of Statistics*, Vol. 1, p. 244.

[4] Choi, K., P. Goldschmidt, and K. Yamashiro (2006). 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), p. 24.

TVAAS 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 considered economically disadvantaged at each school in Tennessee against the school's growth index (the value-added estimate divided by its standard error) for TCAP Mathematics in grades four through eight in 2015. 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: TENNESSEE GROWTH INDEX V. PERCENT TESTED ECONOMICALLY DISADVANTAGED BY SCHOOL

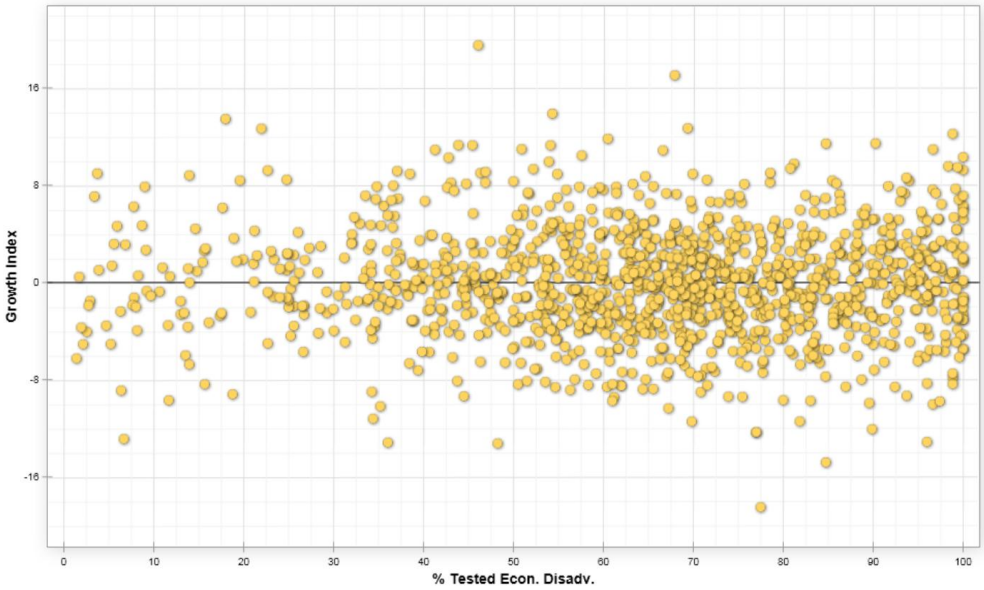
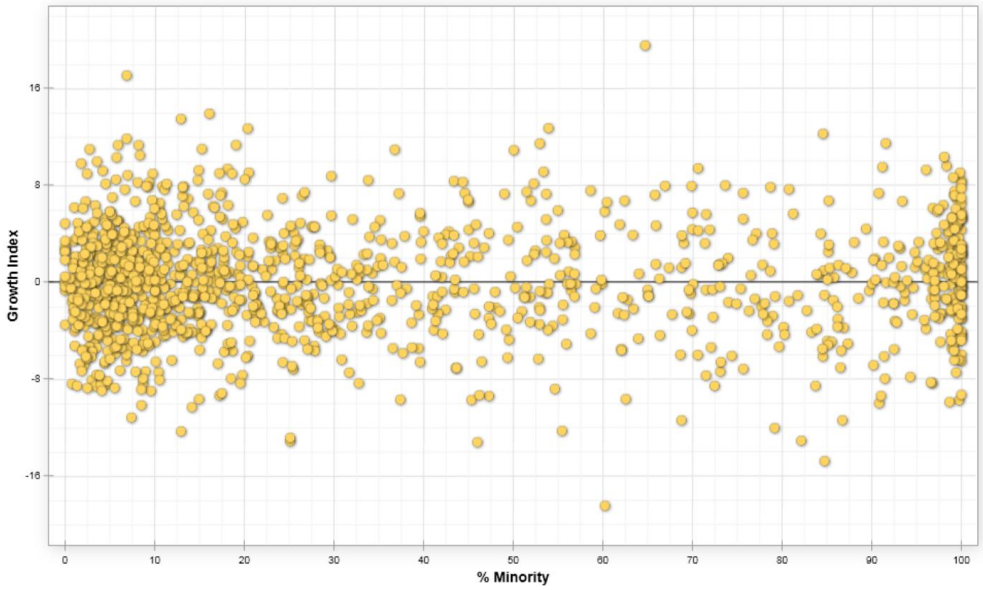


Figure 2 provides similar information for the percentage of minority students. Again, there is essentially no correlation to the growth index.

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



Unintended Policy Consequences of Adjustments

In addition to being statistically unnecessary for an approach like TVAAS, 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.[5] 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.

[5] Mayer, D. P., J.E. Mullens, and M.T. Moore (2000). Monitoring school quality: An indicators report (NCES 2001-030). Washington, DC: National Center for Education Statistics. Source: http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2001030

2 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 progress 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 progress observed in the state in the current year for the subject (for EOC) or subject/grade (for TCAP and K-2 assessments) of interest.

Furthermore, while 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, EOC and K-2 assessment testing scales to measure the growth of both high- and low-achieving students.

In fact, any test that is used in TVAAS analyses must meet three criteria, and the TCAP, EOC and K-2 assessments meet these 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.

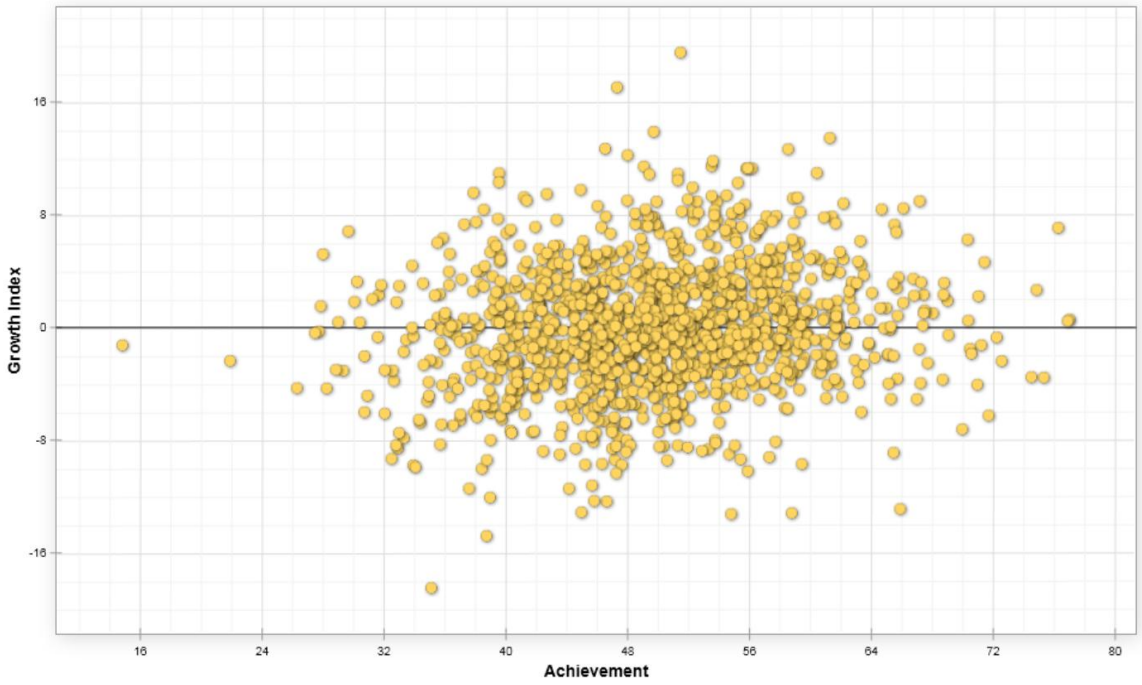
Some educators are concerned about their students who make perfect scores and how that might impact their value-added reporting. In truth, very few students make perfect scores in the same subject from year to year. In 2015, the number of students who made a perfect score in consecutive years for TCAP math was a tiny fraction of a percent—only 0.203%. In the other TCAP subjects, it was even less, ranging from 0.003% in reading to 0.064% in science.

Some educators are concerned about their students who make very low scores and how that may impact their value-added reporting. Because TVAAS is focused on *progress*, rather than *achievement*, and because TVAAS follows the progress of individual students over time, the important thing is that students make about a year's worth of growth. In other words, educators are not disadvantaged by serving low-achieving students who are not yet proficient.

TVAAS 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 Tennessee against its growth index (the value-added estimate divided by its standard error) for TCAP Mathematics in grades four through eight in 2015. 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: TENNESSEE GROWTH INDEX V. AVERAGE ACHIEVEMENT BY SCHOOL



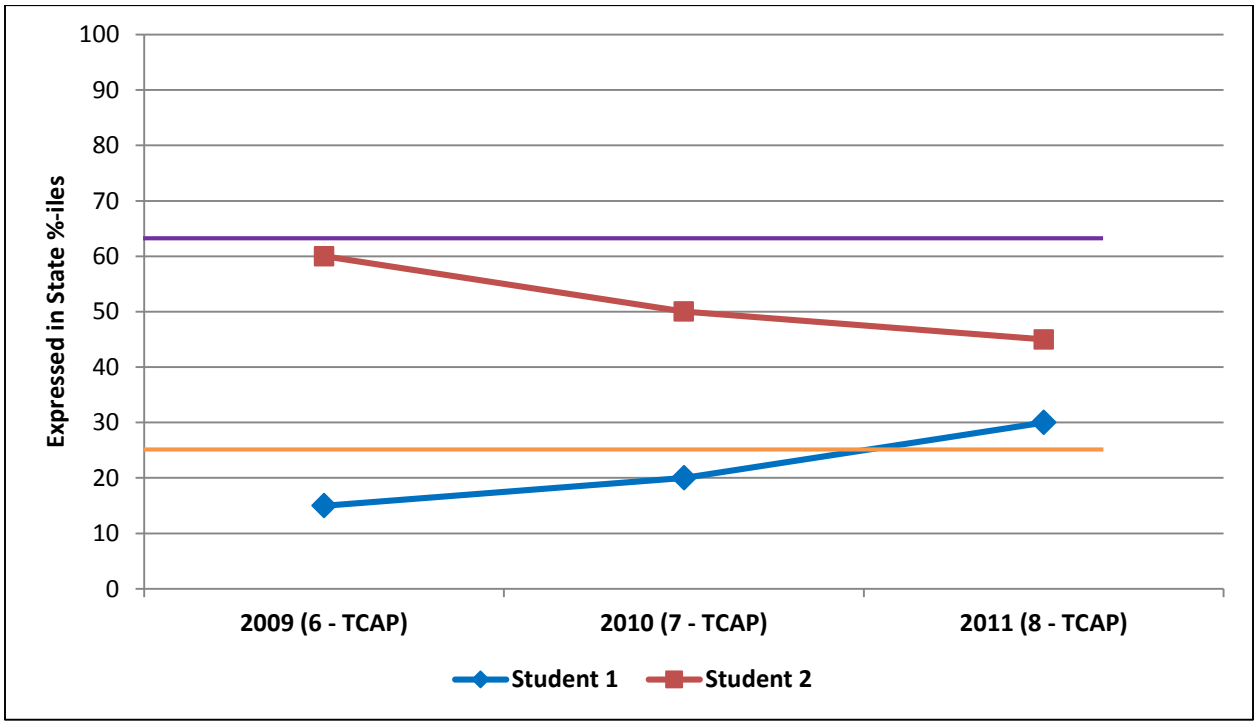
3 TVAAS should always indicate growth if the percentage of students scoring proficient or above increased since last year.

Comparing the percentage of students who score proficient (or above) over time does not account for changes in achievement within performance categories. 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

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 red 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 the 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 TCAP 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.

TVAAS in Practice

TVAAS does not measure students' progress based on the number or percentage of students who tested proficient, accelerated, or advanced, as compared to previous years. TVAAS 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 TVAAS cannot measure the progress of systems and schools with high mobility rates.

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

TVAAS 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 systems 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.

TVAAS 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 system and school growth measures so that highly mobile student populations receive the same level of attention as non-mobile ones.

TVAAS in Practice

For TCAP subjects, all students are included, regardless of their testing history, their number of prior test scores, and which test scores they have. For the K-2 assessments and EOC subjects, 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.

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

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

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 systems 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.

TVAAS 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, TVAAS 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 in the left hand side of Figure 5). However, when there are missing data, each method provides a different result (9.6 vs. 4.0 in the right hand side of Figure 5).

FIGURE 5: EXAMPLE OF STUDENTS WITHOUT AND WITH MISSING DATA

Student	Fourth Grade	Fifth Grade	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	-0.08
8	61.2	64.7	3.5
9	47.3	40.6	-6.7
10	37.8	58.9	5.8
Mean	50.0	55.8	**5.8**
	Diff.	**5.8**	

Student	Fourth Grade	Fifth Grade	Gain
1	51.9		
2	37.9		
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		77.8	
8		64.7	
9	47.3	40.6	-6.7
10	37.8	58.9	5.8
Mean	45.0	54.6	**4.0**
	Diff.	**9.6**	

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

[6] Wright, S. P. (2004). "Advantages of a Multivariate Longitudinal Approach to Educational Value-Added Assessment Without Imputation." Paper presented at National Evaluation Institute, online at http://www.createconference.org/documents/archive/2004/Wright-NEI04.pdf.

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. Furthermore, TVAAS uses much more student data to obtain these relationships in the growth estimates for systems and schools.

TVAAS in Practice

For TCAP subjects, all students are included, regardless of their testing history, their number of prior test scores, and which test scores they have. For K-2 assessments and EOC subjects, all students are included, so 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 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 system and school growth measures so that highly-mobile student populations receive the same level of attention as non-mobile ones.

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

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, TVAAS 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 TVAAS value-added reporting.

TVAAS in Theory

Student test scores are the basic ingredient of all TVAAS analyses. TVAAS is not involved in, and has no control over, test construction. TCAP, EOC and K-2 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:

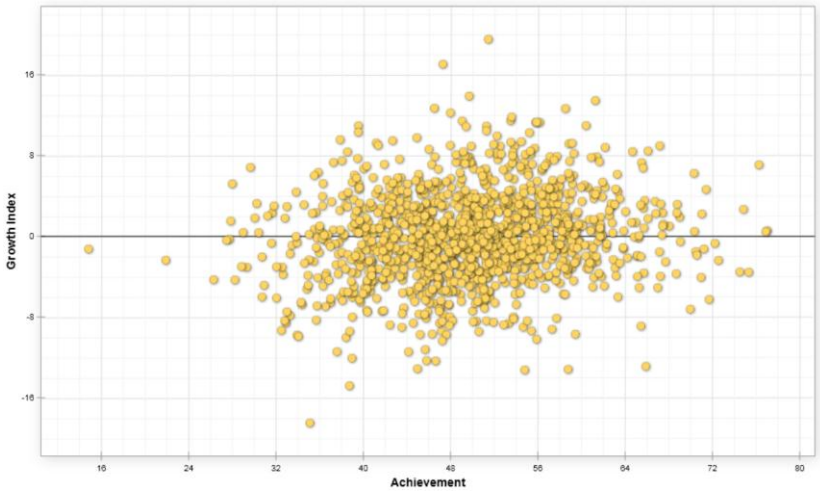
- Must be designed to assess the academic standards.
- Must be reliable and valid (usually related to the number of test questions).
- Must demonstrate sufficient stretch at the extremes.

To date, TCAP, EOC and K-2 assessments have 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 TVAAS estimates are available, SAS verifies that systems, schools and teachers serving both high and low achieving students can show both high and low growth. This process is repeated every year.

TVAAS 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 Tennessee against its growth index (the value-added estimate divided by its standard error) for TCAP Mathematics in grades four through eight in 2015. The graph demonstrates that schools serving both high- and low-achieving students can show both high and low growth, as measured by TVAAS.

FIGURE 6: TENNESSEE GROWTH INDEX V. AVERAGE ACHIEVEMENT BY SCHOOL



7 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. While 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 almost 20 years. Details about the TVAAS models are available in the references below:

- On the **SAS EVAAS Statistical Models upon which Tennessee's reporting is based**: "*SAS® EVAAS® for K-12 Statistical Models*" (2015) available at http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/sas-evaas-k12-statistical-models-107411.pdf.
- 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)

TVAAS in Theory

While TVAAS 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 TVAAS 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 TVAAS modeling. The references below include recent studies by statisticians from the RAND Corporation, a non-profit research organization:

- On the **choice of a complex value-added model**: McCaffrey, D. F., Han, B. and Lockwood, J. R. (2008). "Value-Added Models: Analytic Issues." Prepared for the National Research Council and the National Academy of Education, Board on Testing and Accountability Workshop on Value-Added Modeling, Nov. 13-14, 2008, Washington D.C.
- On the **advantages of the longitudinal, mixed model approach**: Lockwood, J.R. and McCaffrey, D.F. (2007). "Controlling for Individual Heterogeneity in Longitudinal Models, with Applications to Student Achievement." *Electronic Journal of Statistics*, Vol. 1, 223-252.
- On the **insufficiency of simple value-added models**: McCaffrey, D. F., Han, B. and Lockwood, J. R. (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 multivariate response model (MRM) used in value-added analyses is a multivariate, longitudinal, linear mixed model. The MRM 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.
- The univariate response model (URM) used in value-added analyses is conceptually an analysis of covariance (ANCOVA) model. The URM is based on the difference between predicted and observed scores for students. In Tennessee, this is used for the K-2 and EOC reporting.

8 The TVAAS methodology is too complex; a more simple approach to measuring system 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 TVAAS reporting. 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.

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 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.

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

- **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 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.
- **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 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.

- **TVAAS 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.
- **TVAAS 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 TNReady 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:

- On the **choice of a complex value-added model**: McCaffrey, D. F., Han, B. and Lockwood, J. R. (2008). "Value-Added Models: Analytic Issues." Prepared for the National Research Council and the National Academy of Education, Board on Testing and Accountability Workshop on Value-Added Modeling, Nov. 13-14, 2008, Washington D.C.
- On the **advantages of the longitudinal, mixed model approach**: Lockwood, J.R. and McCaffrey, D.F. (2007). "Controlling for Individual Heterogeneity in Longitudinal Models, with Applications to Student Achievement." Electronic Journal of Statistics, Vol. 1, 223-252.
- On the **insufficiency of simple value-added models**: McCaffrey, D. F., Han, B. and Lockwood, J. R. (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

Although the statistical approach is robust and complex, the reports in the TVAAS 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 system or school made more than the expected progress; yellow indicates that students in a system or school made about the expected progress; and red indicates that students in a system 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 TVAAS models by gaining insight in an accessible and non-technical format.

FIGURE 7: SAMPLE TVAAS SYSTEM VALUE-ADDED REPORT

Report: District Value Added **Test:** TCAP
District: Rural County School District **Subject:** Science
Year: 2017

Estimated District Growth Measure

Grade	3	4	5	6	7	8	Growth Measure over Grades Relative to	
Growth Standard		0.0	0.0	0.0	0.0	0.0		
State 3-Yr-Avg							Growth Standard	State
2015 Growth Measure		0.9 G	1.3 G	-1.0 LR	3.1 B	-1.6 R	0.5	
Standard Error		0.6	0.7	0.7	0.6	0.6	0.3	
2016 Growth Measure		4.8 B	2.8 B	7.3 B	0.8 G	4.9 B	4.1	
Standard Error		0.7	0.5	0.5	0.7	0.5	0.3	
2017 Growth Measure		4.1 B	1.7 B	0.9 G	-4.1 R	-7.6 R	-1.0	
Standard Error		0.8	0.6	0.5	0.6	0.5	0.3	
3-Yr-Avg Growth Measure		3.3 B	1.9 B	2.4 B	-0.1 LG	-1.4 R	1.2	
Standard Error		0.4	0.3	0.3	0.4	0.3	0.2	

Estimated District Avg Achievement

Grade	3	4	5	6	7	8
State NCE Average	50.0	50.0	50.0	50.0	50.0	50.0
State 3-Yr-Avg	56.4	55.6	54.9	58.1	57.0	50.3
2014 Avg Achievement	2.3	35.1	37.4	32.8	48.5	42.3
2015 Avg Achievement	47.0	3.2	36.6	36.1	36.7	46.8
2016 Avg Achievement	34.6	51.6	5.6	44.2	37.6	41.5
2017 Avg Achievement	40.5	39.1	53.2	6.6	39.8	29.4

B	Significant evidence that students in the district made more progress than the Growth Standard
G	Moderate evidence that students in the district made more progress than the Growth Standard
LG	Evidence that students in the district made progress similar to the Growth Standard
LR	Moderate evidence that students in the district made less progress than the Growth Standard
R	Significant evidence that students in the district made less progress than the Growth Standard

Achievement results and growth results must be used together to get a complete picture of student learning.
The Growth Standard is met when the student group maintains their position in the state distribution from one year to the next.

9 There is a set distribution of effectiveness levels for districts, schools and teachers.

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

TVAAS in Theory

Due to differences in tests and when students take each test, there are two different value-added models used to measure growth. While similar in concept, the models differ in the precise way that "growth" is determined. A short explanation of each is provided below:

- For TCAP grade 3, EOC, K-2, ACT, PLAN and EXPLORE assessments, the expectation of growth is that students with a district, school, or teacher made the same amount of progress as students with the average district, school, or teacher in the state for that same year/subject/grade. If not all students are taking an assessment in the state, then it may be a subset.
- For TCAP grades 4 – 8, the expectation of growth is that students maintained the same relative position with respect to the statewide student achievement from one year to the next in the same subject area. As an example, if students' achievement was at the 50th NCE in 2014 grade four math, based on the 2014 grade four math statewide distribution of student achievement, and their achievement is at the 50th NCE in 2015 grade five math, based on the 2015 grade five math statewide distribution of student achievement, then their estimated gain is 0.0 NCEs.

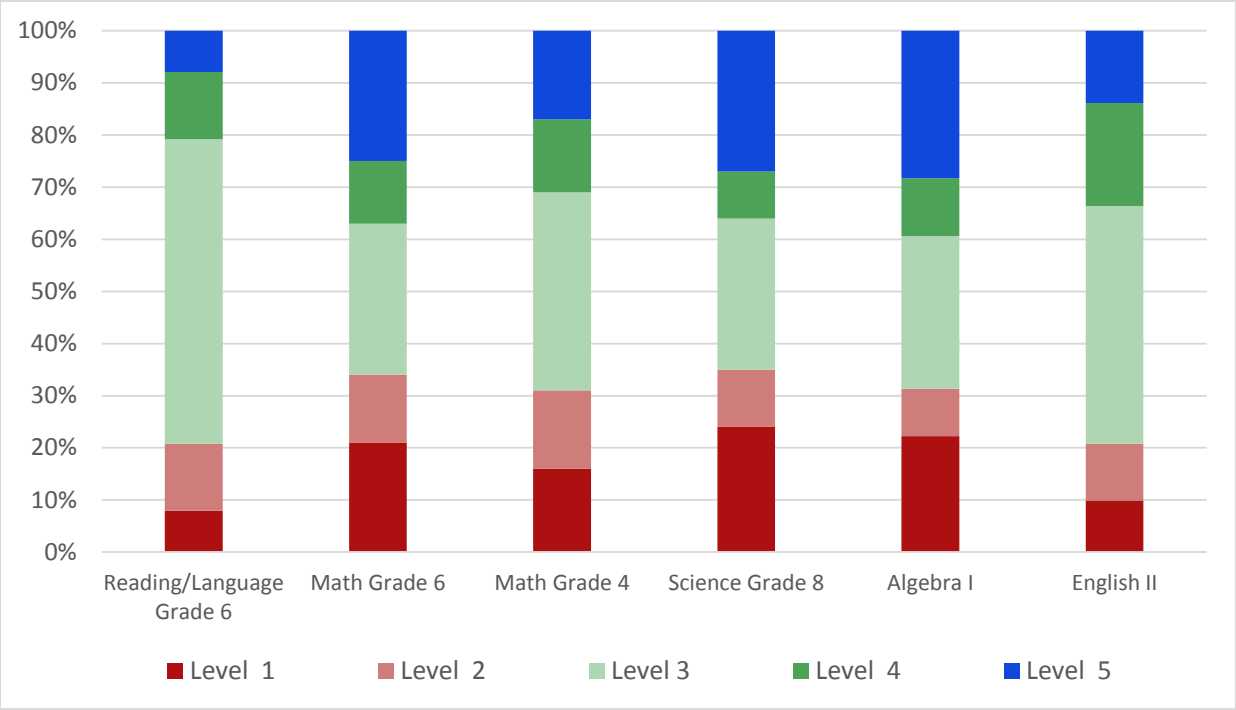
With either approach, the value-added measures tend to be centered on the growth expectation every year, with approximately half of the district/school/teacher estimates above zero and approximately half of the district/school/teacher estimates below zero.

However, it should be noted that there is not a set distribution of the value-added measures and being centered on the growth expectation 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. The actual distribution of levels depends on the extent of measurable differences in student growth among classrooms, schools and districts in a given subject, grade and year. In other words, if students' progress varies more according to their district, school and classroom, then there were be more measurable differences and more districts, schools and teachers in the outer levels (like Level 1 or 5).

TVAAS in Practice

The distribution of effectiveness levels based on single-year 2015 TVAAS measures for teachers for multiple 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 the growth expectation.

FIGURE 8: DISTRIBUTION OF TEACHER EFFECTIVENESS LEVELS BY SUBJECT AND GRADE



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, TVAAS 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.

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 may be more effective with one group than another. 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 reviewed TVAAS value-added estimates from the past two decades and 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 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.

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

TVAAS in Practice

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.

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