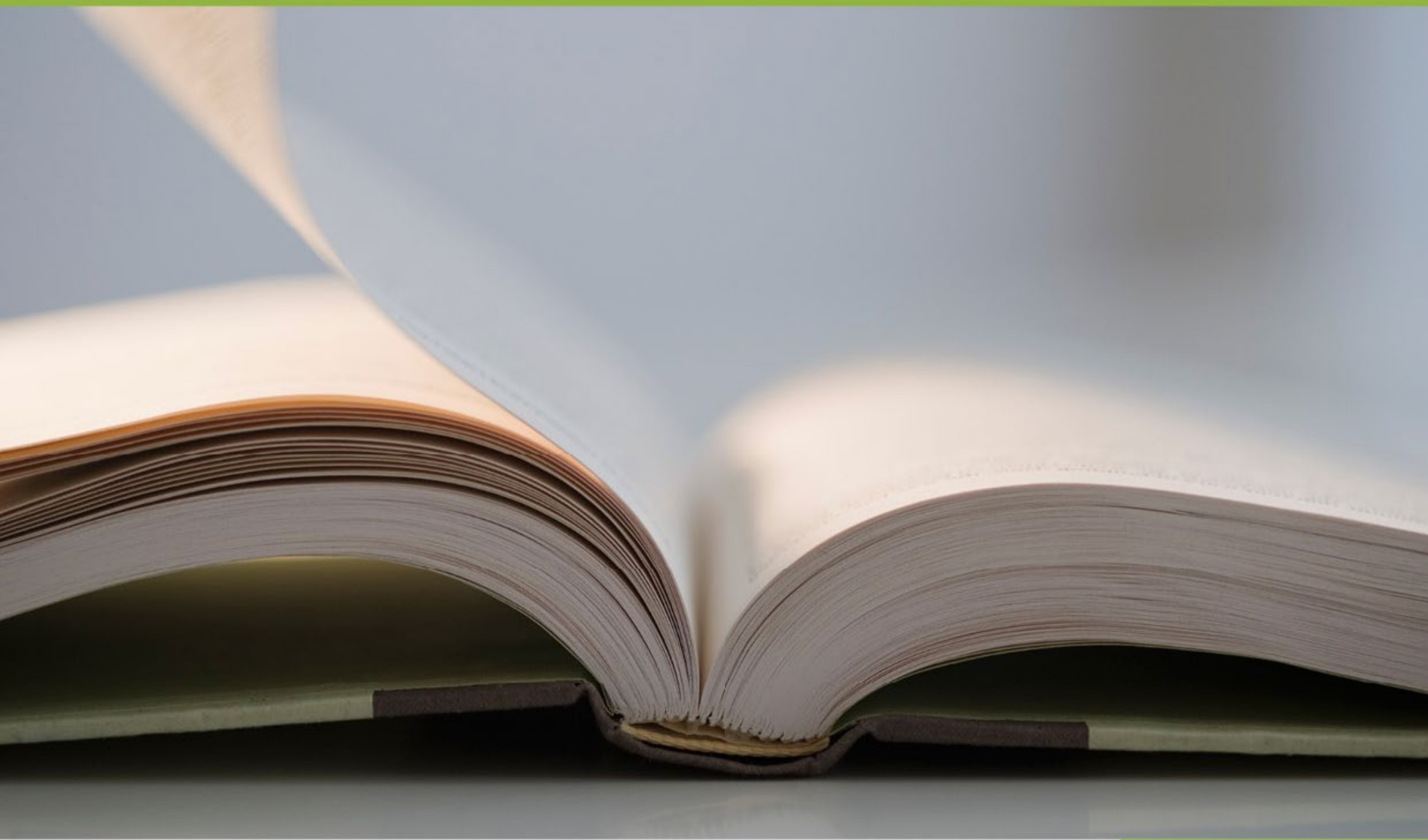
SAS® EVAAS

Topics in Value-Added Modeling

Prepared for the Michigan Department of Education



THE POWER TO KNOW.

Introduction

In 2018, the Michigan Department of Education (MDE) made SAS® EVAAS reporting available to its districts, schools, and teachers. Available through a secure web application, this reporting supports educators with school improvement through both reflective and proactive planning tools.

The core of EVAAS reporting is growth, which measures the change in achievement over time for a group of students. The change is based on student performance on a quality standardized assessment, such as M-STEP or MAP. EVAAS uses a set of growth (or value-added) models that have been available to districts, schools, and teachers in some states since 1993. When first implemented over two decades ago, EVAAS represented a paradigm shift for educators and policymakers to consider both achievement and growth rather than achievement alone. EVAAS reporting provides personalized feedback to districts, schools, and teachers and identifies the more (or less) effective practices in use. This insight can be leveraged to improve the academic experiences of their students.

Conceptually, growth is easy to understand. As stated above, it is simply the change in achievement for a group of students over time. In practice, however, the implementation of a growth model is more complex. The models themselves are highly sophisticated in order to address common questions related to working with assessment data.

The purpose of this document is to address several of these common questions based on the EVAAS growth models.

The information in this document is based on the *typical* EVAAS reporting and includes results from the 2018-19 school year. Due to the pandemic's impact on student learning, the models, interpretation, and results from the 2019-20 and 2020-21 reporting might differ somewhat than what is described here.

Questions related to the student population served by educators

Is student growth correlated with certain demographic variables?

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 EVAAS in the state of Michigan. This is because EVAAS uses multiple subjects, grades, and/or years of testing history for each individual student and does not exclude students who have missing test data. Each student serves as their 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 regard 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 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]

EVAAS 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 might be the most readily apparent evidence to support this belief.

The graph in Figure 1 plots the percentage of tested students who are considered economically disadvantaged at each school in Michigan against the school's growth index (the value-added estimate divided by its standard error) for M-STEP Mathematics in grades 4– 8 in 2019. Regardless of the school's student characteristics, there is typically little or 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.

---

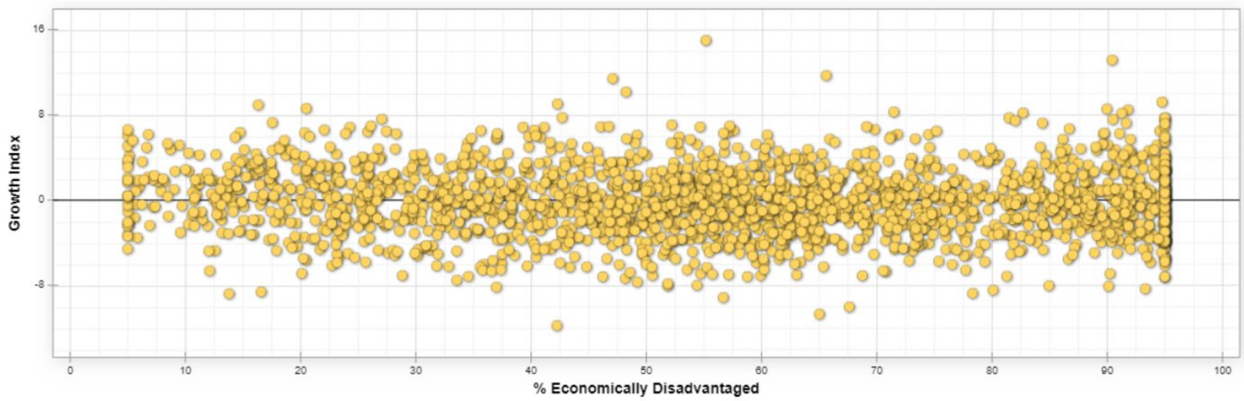
[1] Kevin Carey, "The Real Value of Teachers: Using New Information about Teacher Effectiveness to Close the Achievement Gap," *Thinking K-16* 8, no. 1 (Winter 2004): 27.

[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 1: Michigan Growth Index v. Percent Tested Economically Disadvantaged by School



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

Educators serving either students with histories of higher or lower achievement 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 growth 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 Michigan are designed to follow the growth of individual students over time and estimate whether these students made the average amount of growth observed in the state or population of test-takers in the current year for the assessment of interest.

Furthermore, while the M-STEP assessments are designed to discriminate proficiency from non-proficiency, they, along with interim/benchmark assessments like NWEA MAP, STAR and i-Ready, 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 M-STEP assessment testing scales to measure the growth of both students with histories of high and low achievement.

In fact, any test that is used in EVAAS analyses must meet three criteria, and the M-STEP and interim/benchmark assessments meet these criteria. The tests:

- Must be designed to assess the academic standards.
- Must be sufficiently reliable from one year to the next.
- Must demonstrate sufficient stretch at the extremes to ensure that progress can be measured for both students with histories of high or low achievement

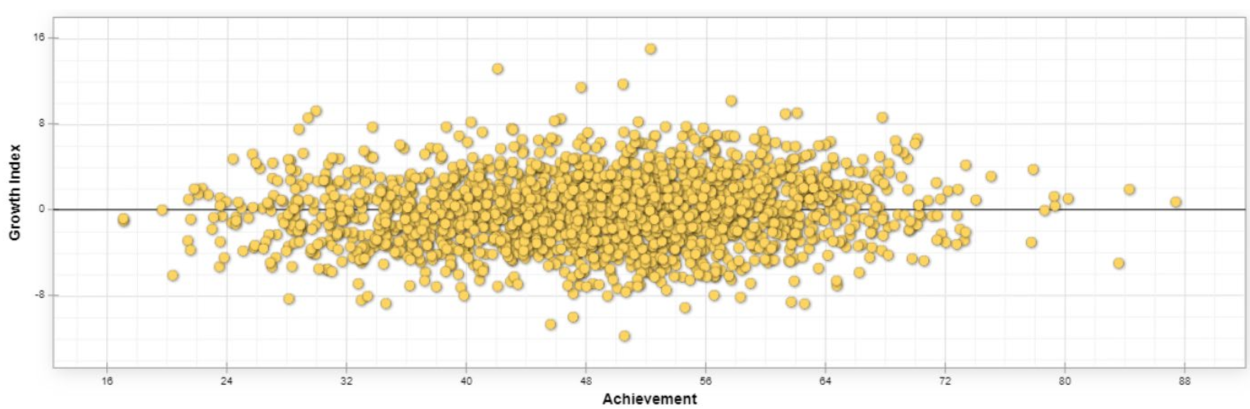
Some educators are concerned about their students who make perfect scores and how that might impact their value-added reporting. In truth, very few students make perfect scores in the same subject from year to year. In 2019, the number of students who made a perfect score in consecutive years for M-STEP Math was a tiny fraction of a percent—only 0.14%. For M-STEP ELA, the percentage was 0.11%.

Some educators are concerned about their students who make very low scores and how that might impact their value-added reporting. However, EVAAS is focused on *growth* rather than *achievement,* and this approach uses multiple years of data, when available, to follow the growth 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. As a conceptual example, if students' average prior achievement was at the 10th NCE (similar to a percentile), the growth model would expect those students' ending achievement to be near the 10th NCE. Likewise, if students' average prior achievement was at the 70th percentile, the growth model would expect those students' exiting achievement to be near the 70th percentile. In other words, educators are not disadvantaged by serving students with a history of low achievement who are not yet proficient.

EVAAS in Practice

Actual data might be the most readily apparent evidence. The graph in Figure 2 plots the average entering achievement for each school in Michigan against its growth index (the value-added estimate divided by its standard error) for M-STEP Mathematics in grades 4–8 in 2019. There is typically little or no correlation between the school's academic achievement and 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 2: Michigan Growth Index v. Average Achievement by School



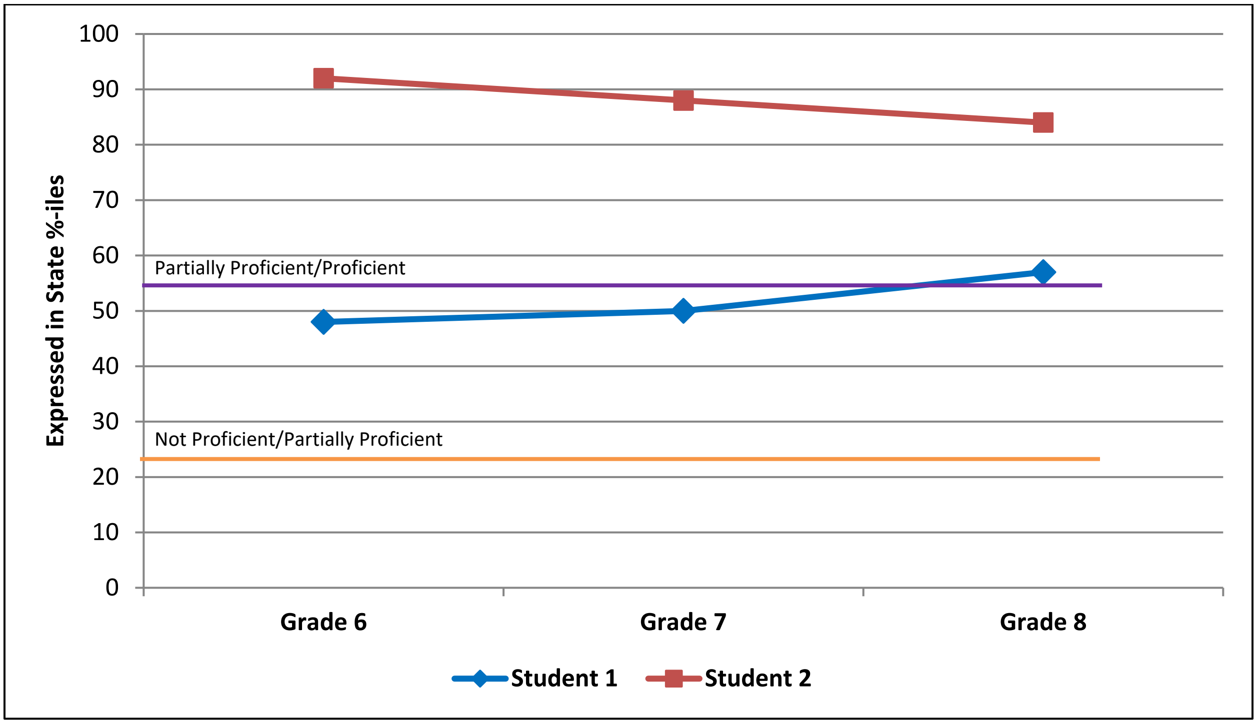
Should EVAAS always indicate growth if the percentage of students scoring Proficient or above increased since last year?

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 growth of individual students over time, regardless of their achievement level, to ensure that all students count. In fact, students' proficiency status is not included in the growth model, as it uses a more precise measure to estimate students' change in achievement over time.

EVAAS in Theory

Imagine the scenario below. The ELA 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 orange and purple lines show the percentile corresponding to the Partially Proficient and Proficient performance levels. 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 Partially Proficient 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 3: Student Testing History in ELA for Student 1 and Student 2



Just by considering the number of students who have scored Proficient, assuming that 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' growth based on the number or percentage of students who tested Proficient or Advanced as compared to previous years. EVAAS instead detects subtle changes in growth within as well as between the performance levels. As a result, educators are recognized when they make growth with students, even if those students did not reach proficiency or if those students maintained their proficiency status.

Can EVAAS 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 have a history of lower achievement , and it is important to include these students to avoid selection bias, which could provide misleading growth estimates to districts and schools that serve students with a history of lower achievement or highly mobile populations of students.[5] Although more simplistic value-added or growth estimates might require that students have the same set of prior test scores or that students have a certain set of required test scores, this often has the result of excluding certain types of students; this would disproportionately affect educators serving those types of students.

EVAAS does not require that students have the same set of prior test scores or all required test scores, 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 4 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 side of Figure 4). However, when there are missing data, each method provides a different result (6.9 vs. 4.6 in the right side of Figure 4).

Figure 4: 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
Mean	50.0	55.8	5.8
	Difference	5.8	

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
Mean	51.2	55.8	6.9
	Difference	4.6	

The problem of missing data is 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

[5] See, for example: David Kernow, "Patterns of Urban Student Mobility and Local School Reform," *Journal of Education for Students Placed at Risk* 1, no. 2 (1996): 147-169.

[6] 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.

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 systems and schools.

EVAAS in Practice

For M-STEP/PSAT 8 in ELA and Mathematics as well as benchmark/local assessments, such as MAP in Mathematics and Reading, all students are included regardless of their testing history, their number of prior test scores, and which test scores they have so long as the students meet the business rules for inclusion in the analysis. For M-STEP Science and Social Studies, PSAT 8/9 for Grade 9, PSAT 10, and SAT, all students are included, as long as they have three prior test scores in any test, grade, and subject and meet the business rules for inclusion in the analysis.

EVAAS reporting is available using Michigan statewide data for the state summative assessments and state-approved benchmark/interim assessments, such as MAP, for districts that opted to submit. As a result, students and their testing history can be tracked as they move within a year (for MAP) and within the state.

Furthermore, regularly excluding highly mobile student populations who tend to be at-risk presents possible problems with educational equity since highly mobile student populations might not otherwise receive the same level of attention as non-mobile students.

Questions related to the tests used in value-added modeling

Is EVAAS reporting reliable or valid since it is based only on state summative assessments?

Educators might be concerned that value-added reporting is limited to the use of state summative tests. Perhaps they feel that the assessments do not correlate well with their curriculum, or that there isn't sufficient stretch to measure growth of very low- or high-achieving students. However, EVAAS estimates use a sophisticated modeling approach to address many of the concerns of using state summative 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 EVAAS analyses. EVAAS is not involved in and has no control over test construction. M-STEP assessments are aligned to the appropriate Michigan grade- and subject-level state 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. The tests:

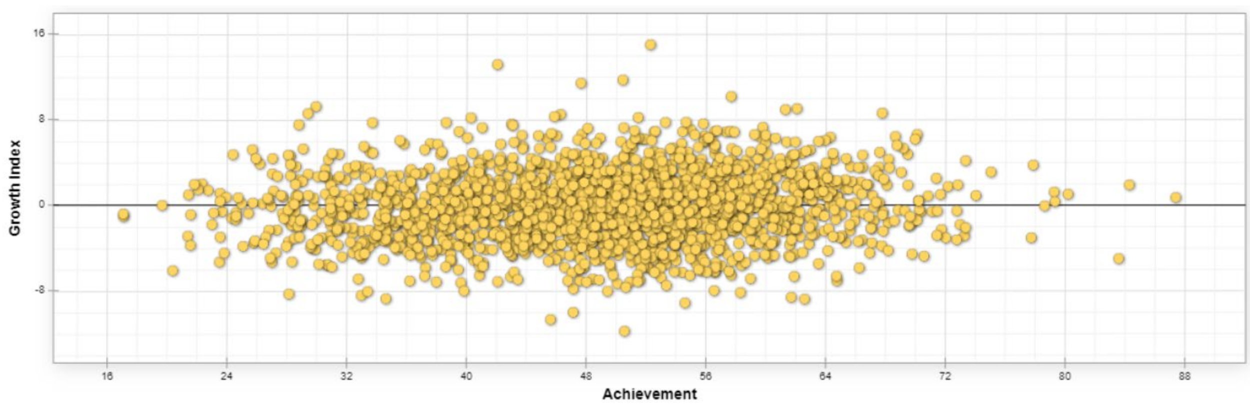
- Must be designed to assess the academic standards.
- Must be sufficiently reliable from one year to the next.
- Must demonstrate sufficient stretch at the extremes to ensure that progress can be measured for both students with histories of high and low achievement.

To date, M-STEP 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 EVAAS estimates are available, SAS verifies that districts and schools serving both students with histories of high and low achievement can show both high and low growth. This process is repeated every year.

EVAAS in Practice

Actual data might be the most readily apparent evidence to support the reliability of the reporting. The graph in Figure 5 plots the average entering achievement for each school in Michigan against its growth index (the value-added estimate divided by its standard error) for M-STEP Mathematics in grades 4–8 in 2019. The graph demonstrates that schools serving both students with histories of high and low achievement can show both high and low growth as measured by EVAAS.

Figure 5: Michigan Growth Index v. Average Achievement by School



Questions related to the value-added modeling approach itself

Has EVAAS methodology been vetted?

EVAAS 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 Michigan is proprietary, the EVAAS methodologies and algorithms are published and have been in the open literature for over 20 years. Details about the EVAAS models are available in the references below:

- On the **SAS EVAAS Statistical Models upon which Michigan'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, Jason (Ed.) Chapters 12-16 in *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* (Thousand Oaks, CA: Corwin, 1997).

EVAAS in Theory

Although EVAAS reporting benefits from a robust modeling approach, this statistical rigor is necessary to provide reliable estimates. More specifically, the 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 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 EVAAS 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, Daniel F., and J.R. Lockwood. 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, DC.
- 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.

EVAAS in Practice

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

- The gain model 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 M-STEP/PSAT 8 reporting in ELA and Mathematics and for the MAP reporting in Mathematics and Reading.
- The predictive model used in value-added analyses is conceptually an analysis of covariance (ANCOVA) model. The predictive model is based on the difference between predicted and observed scores for students. In Michigan, this is used for the M-STEP reporting in Science and Social Studies as well as the PSAT 8/9 in Grade 9, PSAT 10, and SAT reporting.

Why is the EVAAS methodology so complex?

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 student characteristics (prior achievement, demographics, or socioeconomic status) rather than the educator's effectiveness with those students. Such models often unfairly disadvantage educators serving students with histories of low achievement and unfairly advantage educators serving students with histories of high achievement.

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 use all 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/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.
- **EVAAS value-added measures are sophisticated and robust enough to include students with missing data.** Since students with histories of low achievement 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 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 accounts for the inherent uncertainty when providing estimates.
- **EVAAS teacher value-added measures will account for both measures of uncertainty (standard error) and measures of magnitude (effect size) when determining a teacher's effectiveness level.** In addition to accounting for the inherent uncertainty when providing estimates of

student learning, this approach takes into account the practical significance a group of students met, exceeded, or fell short of expected growth.

- **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 different scales. Second, they can continue to provide reporting when the tests change, as was the case when M-STEP replaced MEAP.

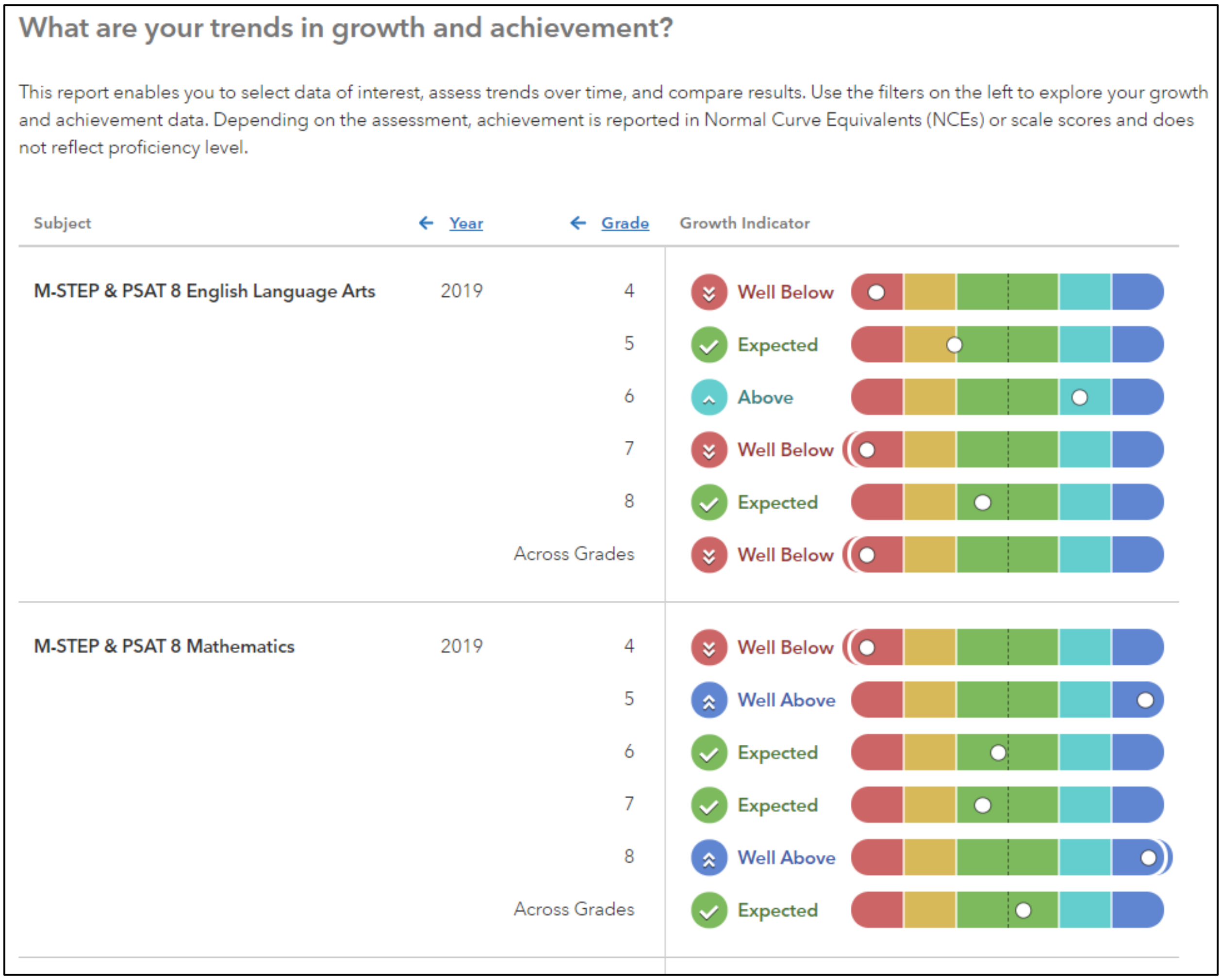
EVAAS statistical models have been validated and vetted by a variety of value-added experts. The references below include recent studies by statisticians from the RAND Corporation, a non-profit research organization:

- On the **choice of a complex value-added model**: McCaffrey, Daniel F. and J.R. Lockwood. 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 DC.
- 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.

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 for quick interpretation: blue indicates that students in a district or school made more than the expected growth; green indicates that students in a district or school made about the expected growth; and red or yellow indicates that students in a district 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 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 EVAAS models by gaining insight in an accessible and non-technical format.

Figure 6: Sample EVAAS District Value-Added Report



Are teacher value-added estimates 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, EVAAS 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 that educators can use to inform a variety of educational and policy decisions. The Michigan Department of Education believes that no single measure should ever be used in isolation to determine high-stakes decisions, and EVAAS is simply one data point.

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 expected. Cohorts of students change every year and teachers might 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 EVAAS 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 can 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 EVAAS: educators and policymakers can rely on the teacher estimates to inform their decisions.

EVAAS in Practice

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