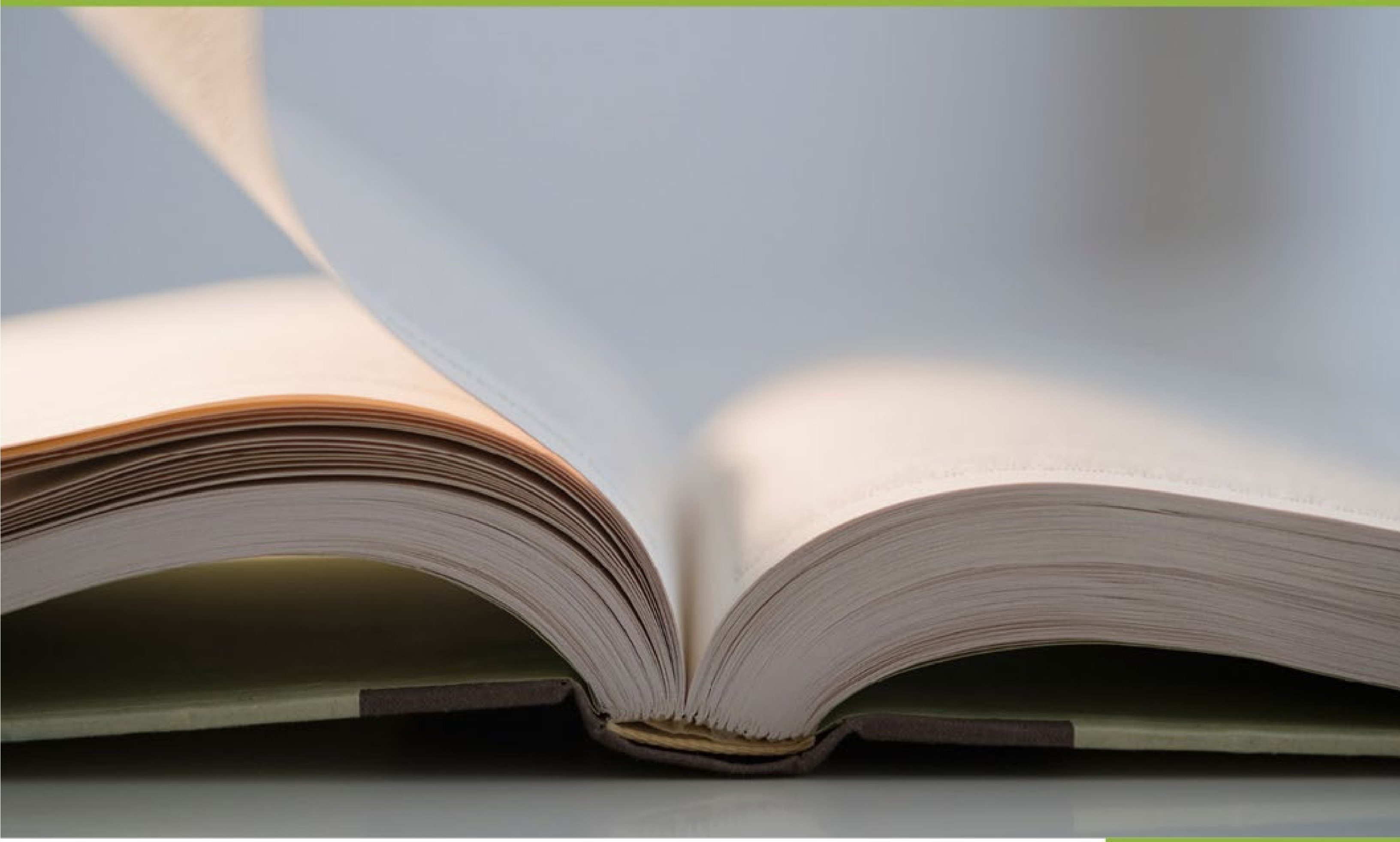
SAS® EVAAS

Statistical Models and Business Rules

Prepared for Michigan Department of Education



THE POWER TO KNOW.

Contents

021122307-838. Copyright © 2022 SAS Institute Inc. Cary, NC, USA. All rights reserved. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

1 Introduction to Michigan's Value-Added Reporting

The term "value-added" refers to a statistical analysis used to measure students' academic growth. Conceptually and as a simple explanation, value-added or growth measures are calculated by comparing the exiting achievement to the entering achievement for a group of students. Although the concept of growth is easy to understand, the implementation of a growth model is more complex.

First, there is not just one growth model; there are multiple growth models depending on the assessment, students included in the analysis, and level of reporting (district, school, or teacher). For each of these models, there are business rules to ensure the growth measures reflect the policies and practices selected by the State of Michigan.

Second, in order to provide reliable growth measures, growth models must overcome non-trivial complexities of working with student assessment data. For example, students do not have the same entering achievement, students do not have the same set of prior test scores, and all assessments have measurement error because they are estimates of student knowledge. EVAAS growth models have been in use and available to educators in states since the early 1990s. These growth models were among the first in the nation to use sophisticated statistical models that addressed these concerns.

Third, the growth measures are relative to students' expected growth, which is in turn determined by the growth that is observed within the actual population of Michigan test-takers in a subject, grade, and year. Interpreting the growth measures in terms of their distance from expected growth provides a more nuanced, and statistically robust, interpretation.

With these complexities in mind, the purpose of this document is to guide you through Michigan's value-added modeling based on the statistical models, business rules, policies, and practices selected by the State of Michigan and currently implemented by EVAAS. This document includes details and decisions in the following areas:

- Conceptual and technical explanations of analytic models
- Definition of expected growth
- Classifying growth into categories for interpretation
- Explanation of district, school, and teacher composites
- Input data
- Business rules

The State of Michigan has provided EVAAS growth measures to Michigan districts, schools, and teachers since 2018. Teacher reporting is made available to any districts that wanted to opt in through the Michigan Data Hub (MiDataHub) project.

These reports are delivered through the EVAAS web application available at http://mi.sas.com. Although the underlying statistical models and business rules supporting these reports are sophisticated and comprehensive, the web reports are designed to be user-friendly and visual so that educators and administrators can quickly identify strengths and opportunities for improvement and then use these insights to inform curricular, instructional, and planning supports.

Of particular note for the 2020-21 reporting is that, in spring 2020, the COVID-19 pandemic required schools to close early and cancel statewide summative assessments. As a result, scores are not available for the M-STEP, PSAT, SAT and end-of-year MAP assessments based on the 2019-20 school year, and the 2020–21 EVAAS reporting does not include these 2019-20 test scores. More details about how this year's EVAAS growth measures were calculated to accommodate the missing year of data are provided in Sections 2.2.4.4 and 2.3.3.1.

2 Statistical Models

2.1 Overview of Statistical Models

The conceptual explanation of value-added reporting is simple: compare students' exiting achievement with their entering achievement over two points in time. In practice, however, measuring student growth is more complex. Students start the school year at different levels of achievement. Some students move around and have missing test scores. Students might have "good" test days or "bad" test days. Tests, standards, and scales change over time. A simple comparison of test scores from one year to the next does not incorporate these complexities. However, a more robust value-added model, such as the one used in Michigan, can account for these complexities and scenarios.

Michigan's value-added models offer the following advantages:

- **The models use multiple subjects and years of data.** This approach minimizes the influence of measurement error inherent in all academic assessments.
- **The models can accommodate students with missing test scores.** This approach means that more students are included in the model and represented in the growth measures. Furthermore, because certain students are more likely to have missing test scores than others, this approach provides less biased growth measures than growth models that cannot accommodate student with missing test scores.
- **The models can accommodate tests on different scales.** This approach gives flexibility to policymakers to change assessments as needed without a disruption in reporting. It permits more tests to receive growth measures, particularly those that are not tested every year.
- **The models can accommodate team teaching or other shared instructional practices.** This approach provides a more accurate and precise reflection of student learning among classrooms.

These advantages provide robust and reliable growth measures to districts, schools, and teachers. This means that the models provide valid estimates of growth given the common challenges of testing data. The models also provide measures of precision along with the individual growth estimates taking into account all of this information.

Furthermore, because this robust modeling approach uses multiple years of test scores for each student and includes students who are missing test scores, EVAAS value-added measures typically have very low correlations with student characteristics. It is not necessary to make *direct* adjustments for student socioeconomic status or demographic flags because each student serves as their own control. In other words, to the extent that background influences persist over time, these influences are already represented in the student's data. As a 2004 study by The Education Trust stated, specifically with regard to the 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.
>
> Source: Carey, Kevin. 2004. "The Real Value of Teachers: Using New Information about Teacher Effectiveness to Close the Achievement Gap." *Thinking K-16* 8(1):27.

In other words, although technically feasible, adjusting for student characteristics in sophisticated modeling approaches is typically not necessary from a statistical perspective, and the value-added reporting in Michigan does not make any direct adjustments for students' socioeconomic/demographic characteristics. Through this approach, the Michigan Department of Education does not provide growth models to educators based on differential expectations for groups of students based on their backgrounds.

Based on Michigan's state assessment program, there are two approaches to providing district, school, and teacher growth measures.

- **Gain model (also known as the multivariate response model or MRM)** is used for tests given in consecutive grades, such as M-STEP Math and ELA in grades 3–7 to provide growth measures in grades 4–7 or MAP Math and Reading in grades 1–8 to provide growth measures in grades 1–8. The gain model is also used to measure growth from grade 7 to 8 with the PSAT 8/9 in grade 8.
- **Predictive model (also known as univariate response model or URM)** is used when a test is given in non-consecutive grades or when performance from previous tests is used to predict performance on another test. This includes M-STEP Science and Social Studies assessments, SAT, and PSAT for grades 9 and 10.

There is another model, which is similar to the predictive model except that it is intended as an instructional tool for educators serving students who have not yet taken an assessment.

- **Projection model** is used for all assessments and provides a probability of obtaining a particular score or higher on a given assessment for individual students.

The following sections provide technical explanations of the models. The online Help within the EVAAS web application is available at https://mi.sas.com, and it provides educator-focused descriptions of the models.

In spring 2020, the COVID-19 pandemic required schools to close early and cancel statewide summative assessments. As a result, statewide scores are not available for Michigan's M-STEP, PSAT 8/9, PSAT 10 and SAT exams based on the 2019-20 school year and the end-of-year MAP assessments. Accommodations for the missing year of data are described for the gain model in Section 2.2.4.4 and for the predictive model in Section 2.3.3.1.

2.2 Gain Model

2.2.1 Overview

The gain model measures growth between two points in time for a group of students; this is the case for tests given in consecutive grades such M-STEP Math and ELA in grades 3–7 to provide growth measures in grades 4–7 or MAP Math and Reading in grades 1–8 to provide growth measures in grades 1–8. The gain model is also used to measure growth from grade 7 to 8 with the PSAT 8/9 in grade 8. **More specifically, the gain model measures the change in relative achievement for a group of students based on the statewide or normed achievement from one point in time to another.** For state summative assessments, growth is typically measured from one year to the next, using the available consecutive grade assessments. For MAP assessments, growth is measured from the beginning of the year to the end of the year within the same grade. Due to suspended assessments in the spring of the

2019-20 school year, the MAP assessments measure growth from the beginning of the year to the middle of the year within the same grade for the 2019-20 reporting.

Expected growth means that students maintained their relative achievement among the population of test-takers, and more details are available in Section 3.

There are three separate analyses for EVAAS reporting based on the gain model: one each for districts, schools, and teachers. The district and school models are essentially the same; they perform well with the large numbers of students characteristic of districts and most schools. The teacher model uses a version adapted to the smaller numbers of students typically found in teachers' classrooms.

In statistical terms, the gain model is known as a linear mixed model and can be further described as a multivariate repeated measures model. These models have been used for value-added analysis for almost three decades, but their use in other industries goes back much further. These models were developed to use in fields with very large longitudinal data sets that tend to have missing data.

Value-added experts consider the gain model to be among one of the most statistically robust and reliable models. The references below include foundational studies by experts from 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 McCaffrey. 2007. "Controlling for Individual Heterogeneity in Longitudinal Models, with Applications to Student Achievement." *Electronic Journal of Statistics* 1:223-252.
- 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.

2.2.2 Why the Gain Model is Needed

A common question is why growth cannot be measured with a simple gain model that measures the difference between the current year's scores and prior year's scores for a group of students. The example in Figure 1 illustrates why a simple approach is problematic.

Assume that 10 students are given a test in two different years with the results shown in Figure 1. The goal is to measure academic growth (gain) from one year to the next. Two simple approaches are to calculate the mean of the differences *or* to calculate the differences of the means. When there is no missing data, these two simple methods provide the same answer (5.8 on the left in Figure 1). When there is missing data, each method provides a different result (6.9 vs. 4.6 on the right in Figure 1).

Figure 1: 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**	**5.8**
Difference between Current and Previous Score Means			**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
Column Mean	**51.2**	**55.8**	**6.9**
Difference between Current and Previous Score Means			**4.6**

A more sophisticated model can account for the missing data and provide a more reliable estimate of the gain. As a brief overview, the gain model uses the correlation between current and previous scores in the non-missing data to estimate means for all previous and current scores as if there were no missing data. It does this without explicitly imputing values for the missing scores. The difference between these two estimated means is an estimate of the average gain for this group of students. In this example, the gain model calculates the estimated difference to be 5.8. Even in a small example such as this, the estimated difference is much closer to the difference with no missing data than either measure obtained by the mean of the differences (6.9) or the difference of the means (4.6). This method of estimation has been shown, on average, to outperform both of the simple methods.[1] This small example only considered two grades and one subject for 10 students. Larger data sets, such as those used in the actual value-added analyses for the state, provide better correlation estimates by having more student data, subjects, and grades. In turn, these provide better estimates of means and gains.

This simple example illustrates the need for a model that will accommodate incomplete data sets, which all student testing sets are. The next few sections provide more technical details about how the gain model calculates student growth.

[1] See, for example, S. Paul Wright, "Advantages of a Multivariate Longitudinal Approach to Educational Value-Added Assessment without Imputation," Paper presented at National Evaluation Institute, 2004. Available online at https://evaas.sas.com/support/EVAAS-AdvantagesOfAMultivariateLongitudinalApproach.pdf.

2.2.3 Common Scale in the Gain Model

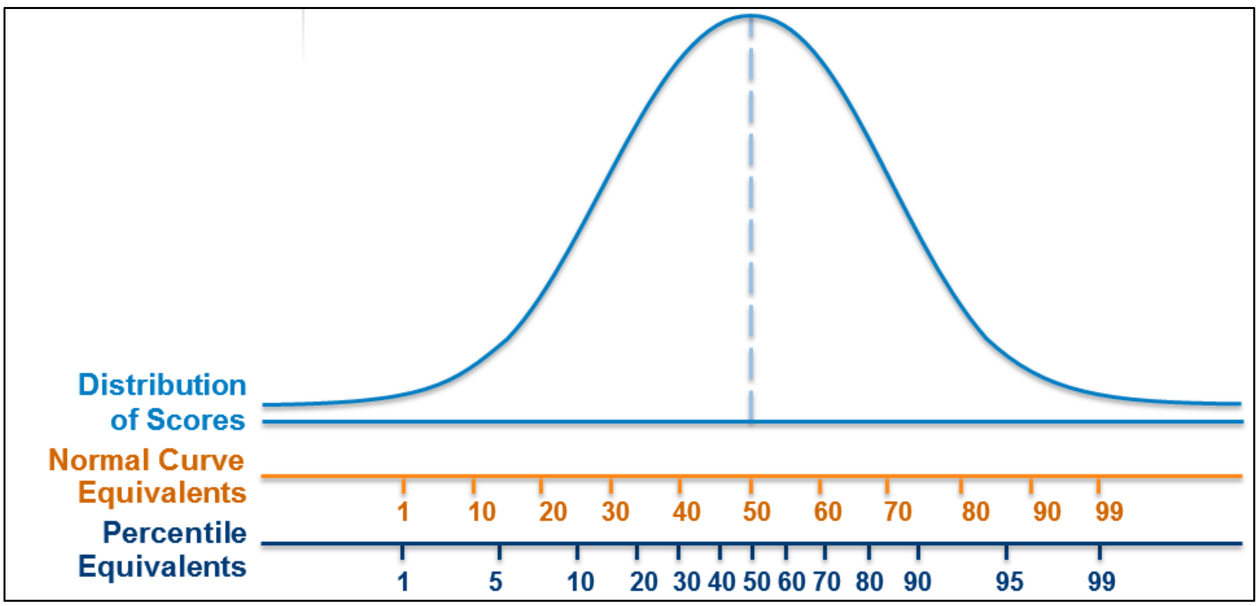
2.2.3.1 Why the Model Uses Normal Curve Equivalents

The gain model estimates academic growth as a "gain," or the difference between two measures of achievement from one point in time to the next. For such a difference to be meaningful, the two measures of achievement (that is, the two tests whose means are being estimated) must measure academic achievement on a common scale. Even for some vertically scaled tests, there can be different growth expectations for students based on their entering achievement. A reliable alternative regardless of whether tests are vertically scaled is to convert scale scores to normal curve equivalents (NCEs).

An NCE distribution is similar to a percentile one. Both distributions provide context as to whether a score is relatively high or low compared to the other scores in the distribution. In fact, NCEs are constructed to be equivalent to percentile ranks at 1, 50 and 99 and to have a mean of 50 and standard deviation of approximately 21.063.

However, NCEs have a critical advantage over percentiles for measuring growth: NCEs are on an equal-interval scale. This means that for NCEs, unlike percentile ranks, the distance between 50 and 60 is the same as the distance between 80 and 90. This difference between the distributions is evident below in Figure 2.

Figure 2: Distribution of Achievement: Scores, NCEs and Percentile Rankings



Furthermore, although percentile ranks are usually truncated below 1 and above 99, NCEs can range below 0 and above 100 to preserve the equal-interval property of the distribution and to avoid truncating the test scale. In a typical year among Michigan's state assessments, the average maximum NCE is approximately 125. Although the gain model does not use truncated values, which could create an artificial floor or ceiling in students' test scores, the web reporting shows NCEs as integers from 1 to 99 for display purposes.

2.2.3.2 Sample Scenario: How to Calculate NCEs in the Gain Model

The NCE distributions used in the gain model are based on a reference distribution of test scores in Michigan for state assessments or a reference distribution of test scores based on national norms for benchmark/interim assessments. This reference distribution is the distribution of scores on a state-mandated test for all students in a given year. By definition, the mean (or average) NCE score for the reference distribution is 50 for each grade and subject. For identifying the other NCEs, the gain model uses a method that does not assume that the underlying scale is normal. This method ensures an equal-interval scale, even if the testing scales are not normally distributed.

Table 1 provides an example of how the gain model converts scale scores to NCEs. The first five columns of the table are based on a subset of Michigan data showing a tabulated distribution of about 45,000 test scores. In a given subject, grade, and year, the tabulation shows, for each given score, the number of students who scored that score ("Frequency") as well as the percentage ("Percent") that frequency represents out of the entire population of test-takers. The table also tabulates the "Cumulative Frequency as the number of students who made that score or lower and its associated percentage ("Cumulative Percent").

The next column, "Percentile Rank," converts each score to a percentile rank. As a sample calculation using the data in Table 1 below, the score of 1477 has a percentile rank of 36.6. The data show that 36.0% of students scored *below* 1477 and 37.2% of students scored *at or below* 1477. To calculate percentile ranks with discrete data, the usual convention is to consider half of the 1.2% reported in the Percent column to be "below" the cumulative percent and "half" above the cumulative percent. To calculate the percentile rank, half of 1.2% (0.6%) is added to 36.0% from Cumulative Percent to give you a percentile rank of 36.6, as shown in the table.

Table 1: Converting Tabulated Test Scores to NCE Values

Score	Frequency	Cumulative Frequency	Percent	Cumulative Percent	Percentile Rank	Z-Score	NCE
1474	1,277	36,632	1.2	33.6	33.0	-0.440	40.74
1475	1,366	37,998	1.3	34.8	34.2	-0.407	41.44
1476	1,299	39,297	1.2	36.0	35.4	-0.373	42.13
1477	1,293	40,590	1.2	37.2	36.6	-0.342	42.80
1478	1,317	41,907	1.2	38.4	37.8	-0.310	43.47
1479	1,299	43,206	1.2	39.6	39.0	-0.279	44.13
1480	1,319	44,525	1.2	40.8	40.2	-0.248	44.79

NCEs are obtained from the percentile ranks using the normal distribution. The table of the standard normal distribution (found in many textbooks[2]) or computer software (for example, a spreadsheet) provides the associated Z-score from a standard normal distribution for any given percentile rank. NCEs

[2] See, for example, the inside front cover of William Mendenhall, Richard L. Scheaffer, and Dennis D. Wackerly, *Mathematical Statistics with Applications* (Boston: Duxbury Press, 1986).

are Z-scores that have been rescaled to have a "percentile-like" scale. As mentioned above, the NCE distribution is scaled so that NCEs exactly match the percentile ranks at 1, 50, and 99. To do this, each Z-score is multiplied by approximately 21.063 (the standard deviation on the NCE scale) and then 50 (the mean on the NCE scale) is added.

With the test scores converted to NCEs, growth is calculated as the difference from one year and grade to the next in the same subject for a group of students. This process is explained in more technical detail in the next section.

2.2.4 Technical Description of the Gain Model

2.2.4.1 Definition of the Linear Mixed Model

As a linear mixed model, the gain model for district, school, and teacher value-added reporting is represented by the following equation in matrix notation:

$$y = X\beta + Zv + \epsilon \tag{1}$$

y (in the growth context) is the $m \times 1$ observation vector containing test scores (usually NCEs) for all students in all academic subjects tested over all grades and years.

X is a known $m \times p$ matrix that allows the inclusion of any fixed effects.

β is an unknown $p \times 1$ vector of fixed effects to be estimated from the data.

Z is a known $m \times q$ matrix that allows the inclusion of random effects.

v is a non-observable $q \times 1$ vector of random effects whose realized values are to be estimated from the data.

ϵ is a non-observable $m \times 1$ random vector variable representing unaccountable random variation.

Both v and ϵ have means of zero, that is, $E(v = 0)$ and $E(\epsilon = 0)$. Their joint variance is given by:

$$Var\begin{bmatrix} v \\ \epsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \tag{2}$$

where R is the $m \times m$ matrix that reflects the amount of variation in and the correlation among the student scores residual to the specific model being fitted to the data, and G is the $q \times q$ variance-covariance matrix that reflects the amount of variation in and the correlation among the random effects. If (v, ϵ) are normally distributed, the joint density of (y, v) is maximized when β has value b and v has value u given by the solution to the following equations, known as Henderson's mixed model equations:[3]

$$\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{bmatrix} \tag{3}$$

[3] McLean, Robert A., William L. Sanders, and Walter W. Stroup (1991). "A Unified Approach to Mixed Linear Models." The American Statistician, Vol. 45, No. 1, pp. 54-64.

Let a generalized inverse of the above coefficient matrix be denoted by

$$\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{bmatrix}^{-} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = C \tag{4}$$

If G and R are known, then some of the properties of a solution for these equations are:

1. Equation (5) below provides the best linear unbiased estimator (BLUE) of the estimable linear function, $K^T \beta$, of the fixed effects. The second equation (6) below represents the variance of that linear function. The standard error of the estimable linear function can be found by taking the square root of this quantity.

$$E(K^T \beta) = K^T b \tag{5}$$

$$Var(K^T b) = (K^T) C_{11} K \tag{6}$$

2. Equation (7) below provides the best linear unbiased predictor (BLUP) of v.

$$E(v|u) = u \tag{7}$$

$$Var(u - v) = C_{22} \tag{8}$$

   where u is unique regardless of the rank of the coefficient matrix.

3. The BLUP of a linear combination of random and fixed effects can be given by equation (9) below provided that $K^T \beta$ is estimable. The variance of this linear combination is given by equation (10).

$$E(K^T \beta + M^T v \,|u) = K^T b + M^T u \tag{9}$$

$$Var(K^T (b - \beta) + M^T (u - v)) = (K^T M^T) C (K^T M^T)^T \tag{10}$$

4. With G and R known, the solution for the fixed effects is equivalent to generalized least squares, and if v and ϵ are multivariate normal, then the solutions for β and v are maximum likelihood.

5. If G and R are not known, then as the estimated G and R approach the true G and R, the solution approaches the maximum likelihood solution.

6. If v and ϵ are not multivariate normal, then the solution to the mixed model equations still provides the maximum correlation between v and u.

2.2.4.2 District and School Models

The district and school gain models do not contain random effects; consequently, the Zv term drops out in the linear mixed model. The X matrix is an incidence matrix (a matrix containing only zeros and ones) with a column representing each interaction of school (in the school model), subject, grade, and year of data. The fixed-effects vector β contains the mean score for each school, subject, grade, and year with each element of β corresponding to a column of X. Since gain models are generally run with each school uniquely defined across districts, there is no need to include districts in the model.

Unlike the case of the usual linear model used for regression and analysis of variance, the elements of ϵ are not independent. Their interdependence is captured by the variance-covariance matrix, which is also known as the R matrix. Specifically, scores belonging to the same student are correlated. If the scores in y are ordered so that scores belonging to the same student are adjacent to one another, then the R matrix is block diagonal with a block, R_i, for each student. Each student's R_i is a subset of the "generic" covariance matrix R_0 that contains a row and column for each subject and grade. Covariances among subjects and grades are assumed to be the same for all years (technically, all cohorts), but otherwise the R_0 matrix is unstructured. Each student's R_i contains only those rows and columns from R_0 that match the subjects and grades for which the student has test scores. In this way, the gain model is able to use all available scores from each student.

Algebraically, the district gain model is represented as:

$$y_{ijkld} = \mu_{jkld} + \epsilon_{ijkld} \tag{11}$$

where y_{ijkld} represents the test score for the i^{th} student in the j^{th} subject in the k^{th} grade during the l^{th} year in the d^{th} district. μ_{jkld} is the estimated mean score for this particular district, subject, grade, and year. ϵ_{ijkld} is the random deviation of the i^{th} student's score from the district mean.

The school gain model is represented as:

$$y_{ijkls} = \mu_{jkls} + \epsilon_{ijkls} \tag{12}$$

This is the same as the district analysis with the addition of the subscript s representing s^{th} school.

The gain model uses multiple years of student testing data to estimate the covariances that can be found in the matrix R_0. This estimation of covariances is done within each level of analyses and can result in slightly different values within each analysis.

Solving the mixed model equations for the district or school gain model produces a vector b that contains the estimated mean score for each school (in the school model), subject, grade, and year. To obtain a value-added measure of average student growth, a series of computations can be done using the students from a school in a particular year and their prior and current testing data. The model produces means in each subject, grade, and year that can be used to calculate differences in order to obtain gains. Because students might change schools from one year to the next (in particular when transitioning from elementary to middle school, for example), the estimated mean score for the prior year/grade uses students who existed in the current year of that school. Therefore, mobility is taken into account within the model. Growth of students is computed using all students in each school including those that might have moved buildings from one year to the next.

The computation for obtaining a growth measure can be thought of as a linear combination of fixed effects from the model. The best linear unbiased estimate for this linear combination is given by equation (5). The growth measures are reported along with standard errors, and these can be obtained by taking the square root of equation (6) as described above.

2.2.4.3 Teacher Model

The teacher estimates use a more conservative statistical process to lessen the likelihood of misclassifying teachers. Each teacher's growth measure is assumed to be equal to the state average (or the average of the nationally representative sample for benchmark/interim assessments) in a specific year, subject, and grade until the weight of evidence pulls them either above or below that state average. The model also accounts for the percentage of instructional responsibility the teacher has for each student during the course of each school year. Furthermore, the teacher model is "layered," which means that:

- Students' performance with both their current and previous teacher effects are incorporated.
- For each school year, the teacher estimates are based students' testing data collected over multiple previous years.

Each element of the statistical model for teacher value-added modeling provides an additional level of protection against misclassifying each teacher estimate.

To allow for the possibility of many teachers with relatively few students per teacher, the gain model enters teachers as random effects via the Z matrix in the linear mixed model. The X matrix contains a column for each subject, grade, and year, and the b vector contains an estimated state mean score for each subject, grade, and year. The Z matrix contains a column for each subject, grade, year, and teacher, and the u vector contains an estimated teacher effect for each subject, grade, year, and teacher. The R matrix is as described above for the district or school model. The G matrix contains teacher variance components with a separate unique variance component for each subject, grade, and year. To allow for the possibility that a teacher might be very effective in one subject and very ineffective in another, the G matrix is constrained to be a diagonal matrix. Consequently, the G matrix is a block diagonal matrix with a block for each subject/grade/year. Each block has the form $\sigma^2{}_{jkl}I$ where $\sigma^2{}_{jkl}$ is the teacher variance component for the j^{th} subject in the k^{th} grade in the l^{th} year, and I is an identity matrix.

Algebraically, the teacher model is represented as:

$$y_{ijkl} = \mu_{jkl} + \left(\sum_{k^* \leq k} \sum_{t=1}^{T_{ijk^*l^*}} w_{ijk^*l^*t} \times \tau_{jk^*l^*t} \right) + \epsilon_{ijkl} \tag{13}$$

y_{ijkl} is the test score for the i^{th} student in the j^{th} subject in the k^{th}grade in the l^{th} year. $\tau_{jk^*l^*t}$ is the teacher effect of the t^{th} teacher in the j^{th} subject in grade k^* in year l^*. The complexity of the parenthesized term containing the teacher effects is due to two factors. First, in any given subject, grade, and year, a student might have more than one teacher. The inner (rightmost) summation is over all the teachers of the i^{th} student in a particular subject, grade, and year, denoted by $T_{ijk^*l^*}$. $\tau_{jk^*l^*t}$ is the effect of the t^{th} teacher. $w_{ijk^*l^*t}$ is the fraction of the i^{th} student's instructional time claimed by the t^{th} teacher. Second, as mentioned above, this model allows teacher effects to accumulate over time. The outer (leftmost) summation accumulates teacher effects not only for the current (subscripts k and l) but also over previous grades and years (subscripts k^* and l^*) in the same subject. Because of this accumulation of teacher effects, this type of model is often called the "layered" model.

In contrast to the model for many district and school estimates, the value-added estimates for teachers are not calculated by taking differences between estimated mean scores to obtain mean gains. Rather, this teacher model produces teacher "effects" (in the u vector of the linear mixed model). It also produces state-level mean scores (for each year, subject, and grade) in the fixed-effects vector b. Because of the way the X and Z matrices are encoded, in particular because of the "layering" in Z, teacher gains can be estimated by adding the teacher effect to the state mean gain. That is, the interpretation of a teacher effect in this teacher model is as a gain expressed as a deviation from the average gain for the state in a given year, subject, and grade.

Table 2 illustrates how the Z matrix is encoded for three students who have three different scenarios of teachers during grades 3, 4, and 5 in two subjects, Math (M) and Reading (R). Teachers are identified by the letters A–F.

Tommy's teachers represent the conventional scenario. Tommy is taught by a single teacher in both subjects each year (teachers A, C, and E in grades 3, 4, and 5, respectively). Notice that in Tommy's Z matrix rows for grade 4 there are ones (representing the presence of a teacher effect) not only for fourth-grade teacher C but also for third-grade teacher A. This is how the "layering" is encoded. Similarly, in the grade 5 rows, there are ones for grade 5 teacher E, grade 4 teacher C, and grade 3 teacher A.

Susan is taught by two different teachers in grade 3: teacher A for Math and teacher B for Reading. In grade 4, Susan had teacher C for Reading. For some reason, in grade 4 no teacher claimed Susan for Math even though Susan had a grade 4 Math test score. This score can still be included in the analysis by entering zeros into the Susan's Z matrix rows for grade 4 Math. In grade 5, however, Susan had no test score in Reading. This row is completely omitted from the Z matrix. There will always be a Z matrix row corresponding to each test score in the y vector. Since Susan has no entry in y for grade 5 Reading, there can be no corresponding row in Z.

Eric's scenario illustrates team teaching. In grade 3 Reading, Eric received an equal amount of instruction from teachers A and B. The entries in the Z matrix indicate each teacher's contribution, 0.5 for each teacher. In grade 5 Math, however, Eric was taught by both teachers E and F, but they did not make an equal contribution. Teacher E claimed 80% responsibility, and teacher F claimed 20%.

Because teacher effects are treated as random effects in this approach, their estimates are obtained by shrinkage estimation, which is technically known as best linear unbiased prediction or as empirical Bayesian estimation. This means that *a priori* a teacher is considered "average" (with a teacher effect of zero) until there is sufficient student data to indicate otherwise. This method of estimation protects against false positives (teachers incorrectly evaluated as most effective or least effective), particularly in the case of teachers with few students.

Table 2: Encoding the Z Matrix

			Teachers											
			Third Grade				Fourth Grade				Fifth Grade			
			A		B		C		D		E		F	
Student	Grade	Subjects	M	R	M	R	M	R	M	R	M	R	M	R
Tommy	**3**	**M**	1	0	0	0	0	0	0	0	0	0	0	0
		R	0	1	0	0	0	0	0	0	0	0	0	0
	4	**M**	1	0	0	0	1	0	0	0	0	0	0	0
		R	0	1	0	0	0	1	0	0	0	0	0	0
	5	**M**	1	0	0	0	1	0	0	0	1	0	0	0
		R	0	1	0	0	0	1	0	0	0	1	0	0
Susan	**3**	**M**	1	0	0	0	0	0	0	0	0	0	0	0
		R	0	0	0	1	0	0	0	0	0	0	0	0
	4	**M**	1	0	0	0	0	0	0	0	0	0	0	0
		R	0	0	0	1	0	1	0	0	0	0	0	0
	5	**M**	1	0	0	0	0	0	0	0	0	0	1	0
Eric	**3**	**M**	1	0	0	0	0	0	0	0	0	0	0	0
		R	0	0.5	0	0.5	0	0	0	0	0	0	0	0
	4	**M**	1	0	0	0	0	0	1	0	0	0	0	0
		R	0	0.5	0	0.5	0	0	0	1	0	0	0	0
	5	**M**	1	0	0	0	0	0	1	0	0.8	0	0.2	0
		R	0	0.5	0	0.5	0	0	0	1	0	0	0	1

From the computational perspective, the teacher gain can be defined as a linear combination of both fixed effects and random effects and is estimated by the model using equation (9). The variance and standard error can be found using equation (10).

2.2.4.4 Accommodations to the Gain Model for Missing 2019-20 Data Due to the Pandemic

In spring 2020, the COVID-19 pandemic required schools to close early and cancel statewide summative assessments. As a result, scores are not available for Michigan's M-STEP assessments based on the 2019-20 school year, and it is not possible to measure growth on the M-STEP assessments from the 2018-19 to the 2019-20 school years or from the 2019-20 to the 2020-21 school years. For the gain model based on M-STEP Math and ELA and PSAT 8/9 in grade 8, the 2020-21 reporting measures growth from the 2018-19 school year to the 2020-21 school year. Because interim/benchmark assessments are administered at several points throughout the school year, the gain model for MAP growth in the same way it has in previous years, except that 2019-20 measures growth from the beginning of year (BOY) to the middle of year (MOY) rather than BOY to end of year (EOY).

From a technical perspective, the gain model for M-STEP in Math and ELA is essentially the same as it has been in previous years except that growth is measured over two years rather than one year. However, the interpretation of these growth measures changes slightly in two notable ways.

First, because the models provide two-year growth measures, the growth measure for grades where students transition from one school to another will then include growth from the feeder school(s) as well as the receiver school. For example, a middle school with grades 6–8 could receive a growth measure for sixth grade based on the students' growth in sixth grade as well as their growth from the feeder elementary school(s) in fifth grade.

In other words, it is not possible to parse out the individual contribution of the middle school in sixth grade apart from those from the elementary school(s) in fifth grade because of the missing year of test scores. For the district-level growth measures and for the non-transition grades, the two-year growth measures are still solely representative of growth within the specific district and the non-transition grades for the school are still solely representative of growth within the specific school.

Second, at a particular school, the growth of certain groups of students are not represented in the two-year measures as they would be in two one-year growth measures. For example, it is not possible to measure the growth of grade 4 students this year because there is no grade 3 data from last year and no statewide assessment to use from grade 2 in 2019. Similarly, it is not possible to report grade 8 growth from last year because there is no exiting achievement for these students in their last year at the school.

Despite these differences, the conceptual explanation of the 2020-21 growth measures is the same as it has always been: these growth measures compare students' exiting achievement with their entering achievement over two points in time.

2.3 Predictive Model

2.3.1 Overview

Tests that are not given in consecutive grades require a different modeling approach from the gain model. The predictive model is used for such assessments in Michigan. **The predictive model is a regression-based model where growth is a function of the difference between students' expected**

scores with their actual scores. Expected growth is met when students with a district or school made the same amount of growth as students with the average district or school.

There are two separate analyses for EVAAS reporting based on the predictive model: one each for districts and schools. There is no teacher model because the assessments that use the predictive model do not receive Teacher reports. The district and school models are essentially the same.

Regression models are used in virtually every field of study, and their intent is to identify relationships between two or more variables. When it comes to measuring growth, regression models identify the relationship between prior test performance and actual test performance for a given course. In more technical terms, the predictive model is known as the univariate response model (URM), a linear mixed model and, more specifically, an analysis of covariance (ANCOVA) model.

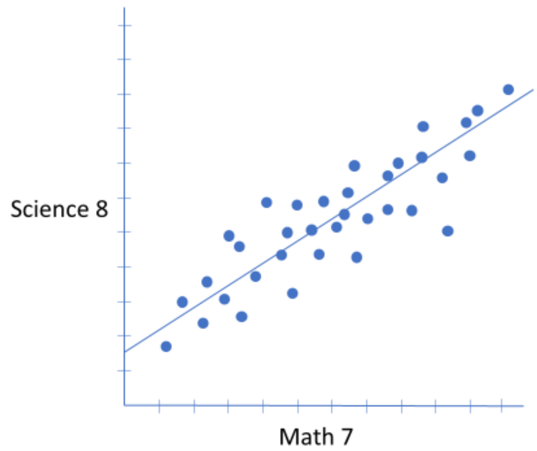
The key advantages of the predictive model can be summarized as follows:

- It minimizes the influence of measurement error and increases the precision of predictions by using multiple prior test scores as predictors for each student.
- It does not require students to have all predictors or the same set of predictors as long as a student has at least three prior test scores as predictors of the response variable in any subject and grade.
- It allows educators to benefit from all tests, even when tests are on differing scales.

2.3.2 Conceptual Explanation

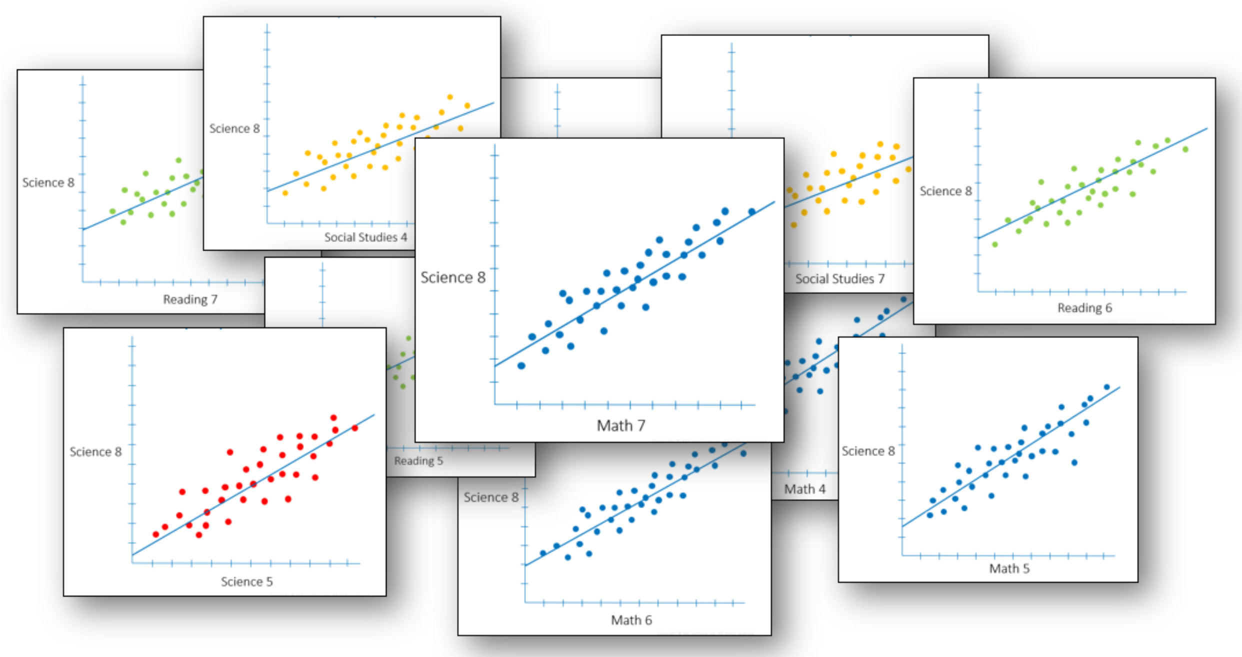
As mentioned above, the predictive model is ideal for assessments given in non-consecutive grades, such as M-STEP Science and Social Studies assessments and SAT. It is also used for PSAT in grades 9 and 10. Consider all students who tested in M-STEP Science in grade 8 in a given year. The gain model is not possible since there isn't a science test in the immediate prior grade. However, these students might have a number of prior test scores in M-STEP Math and ELA in grades 3–7 and M-STEP Social Studies in grade 5. These prior test scores have a relationship with M-STEP Science in grade 8, meaning that how students performed on these tests can predict how the students perform on M-STEP Science in grade 8. The growth model does not assume what the predictive relationship will be; instead, the actual relationships observed by the data define the relationships. This is shown in Figure 3 below where each dot represents a student's prior score on M-STEP Math grade 7 plotted with their score on M-STEP Science grade 8. The best-fit line indicates how students with a certain prior score on M-STEP Math grade 7 tend to score, on average, on M-STEP Science grade 8. This illustration is based on one prior test; the predictive model uses many prior test scores from different subjects and grades.

Figure 3: Test Scores from One Assessment Have a Predictive Relationship to Test Scores from Another Assessment



Some subjects and grades will have a greater relationship to M-STEP Science in grade 8 than others; however, the other subjects and grades still have a predictive relationship. For example, prior Math scores might have a stronger predictive relationship to M-STEP Science in grade 8 than prior ELA scores, but how a student performs on the M-STEP ELA assessment typically provides an idea of how we might expect a student to perform on average on M-STEP Science assessment. This is shown in Figure 4 below, where there are a number of different assessments that have a predictive relationship with M-STEP Science in grade 8. All of these relationships are considered together in the predictive model with some assessments weighted more heavily than others.

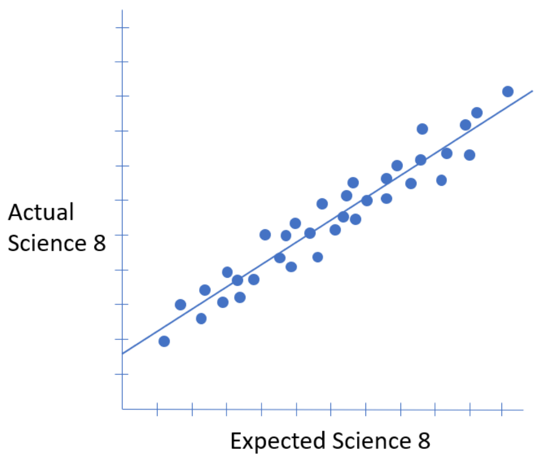
Figure 4: Relationships Observed in the Statewide Data Inform the Predictive Model



Note that the prior test scores do not need to be on the same scale as the assessment being measured for student growth. Just as height (reported in inches) and weight (reported in pounds) can predict a child's age (reported in years), the growth model can use test scores from different scales to find the predictive relationship.

Each student receives an expected score based on their own prior testing history. In practical terms, the expected score represents the student's entering achievement because it is based on all prior testing information to date. Figure 5 below shows the relationship between expected and actual scores for a group of students.

Figure 5: Relationship Expected Score and Actual Score for Selected Subject and Grade



The expected scores can be aggregated to a specific district or school and then compared to the students' actual scores. In other words, the growth measure is a function of the difference between the exiting achievement (or average actual score) and the entering achievement (or average expected score) for a group of students. Unlike the gain model, the actual score and expected score are reported in the scaling units of the test rather than NCEs.

2.3.3 Technical Description of the District and School Models

The predictive model has similar approaches for districts and schools. The approach is described briefly below, with more details following.

- The score to be predicted serves as the response variable (y, the dependent variable).
- The covariates (x's, predictor variables, explanatory variables, independent variables) are scores on tests the student has taken in previous years from the response variable.
- There is a categorical variable (class variable, grouping variable) to identify the district or school from whom the student received instruction in the subject, grade, and year of the response variable (y).

Algebraically, the model can be represented as follows for the i^{th} student.

$$y_i = \mu_y + \alpha_j + \beta_1(x_{i1} - \mu_1) + \beta_2(x_{i2} - \mu_2) + \cdots + \epsilon_i \qquad (14)$$

The μ terms are means for the response and the predictor variables. α_j is the district/school effect for the j^{th} district/school. The β terms are regression coefficients. Predictions to the response variable are made by using this equation with estimates for the unknown parameters (μs, βs, and sometimes α_j). The parameter estimates (denoted with "hats," e.g., $\hat{\mu}$, $\hat{\beta}$) are obtained using all students that have an observed value for the specific response and have three predictor scores. The resulting prediction equation for the i^{th} student is as follows:

$$\hat{y}_i = \hat{\mu}_y + \hat{\beta}_1(x_{i1} - \hat{\mu}_1) + \hat{\beta}_2(x_{i2} - \hat{\mu}_2) + \cdots \quad (15)$$

Two difficulties must be addressed in order to implement the estimation and use of this model. First, not all students will have the same set of predictor variables due to missing test scores. Second, because the predictive model is an ANCOVA model with school as a random effect, the regression coefficients are pooled within group (district or school). The strategy for dealing with missing predictors is to estimate the joint covariance matrix (call it C) of the response and the predictors. Let C be partitioned into response (y) and predictor (x) partitions, that is,

$$C = \begin{bmatrix} c_{yy} & c_{yx} \\ c_{xy} & C_{xx} \end{bmatrix} \quad (16)$$

Note that *C* in equation (16) is not the same as *C* in equation (4). This matrix is estimated using the EM (expectation maximization) algorithm for estimating covariance matrices in the presence of missing data available in SAS/STAT® (although no imputation is actually used). It should also be noted that, because this model is an ANCOVA model, *C* is a pooled-within group (district or school) covariance matrix. This is accomplished by providing scores to the EM algorithm that are centered around group means (i.e., the group means are subtracted from the scores) rather than around grand means. Obtaining *C* is an iterative process since group means are estimated within the EM algorithm to accommodate missing data. Once new group means are obtained, another set of scores is fed into the EM algorithm again until *C* converges. This overall iterative EM algorithm is what accommodates the two difficulties mentioned above. The estimation only includes students who had a test score for the response variable in the most recent year *and* who had at least three predictor variables are included in the estimation. Given such a matrix, the vector of estimated regression coefficients for the projection equation (15) can be obtained as:

$$\hat{\beta} = C_{xx}^{-1} c_{xy} \quad (17)$$

This allows one to use whichever predictors a student has to get that student's expected y-value ($\hat{y}_i$). Specifically, the C_{xx} matrix used to obtain the regression coefficients *for a particular student* is that subset of the overall C matrix that corresponds to the set of predictors for which this student has scores.

The prediction equation also requires estimated mean scores for the response and for each predictor (the $\hat{\mu}$ terms in the prediction equation). These are not simply the grand mean scores. It can be shown that in an ANCOVA if one imposes the restriction that the estimated "group" effects should sum to zero (that is, the effect for the "average" district or school is zero), then the appropriate means are the means of the group means. The group-level means are obtained from the EM algorithm mentioned above, which accounts for missing data. The overall means ($\hat{\mu}$ terms) are then obtained as the simple average of the group-level means.

Once the parameter estimates for the prediction equation have been obtained, predictions can be made for any student with any set of predictor values as long as that student has a minimum of three prior test scores. This is to avoid bias due to measurement error in the predictors. For the 2020-21 reporting, expected scores for M-STEP Science and Social Studies in grade 5 are based on only two predictors due to missing data from the 2019-20 school year

$$\hat{y}_i = \hat{\mu}_y + \hat{\beta}_1(x_{i1} - \hat{\mu}_1) + \hat{\beta}_2(x_{i2} - \hat{\mu}_2) + \cdots \tag{18}$$

The $\hat{y}_i$ term is nothing more than a composite of all the student's past scores. It is a one-number summary of the student's level of achievement prior to the current year, and this term is called the expected score or entering achievement in the web reporting. The different prior test scores making up this composite are given different weights (by the regression coefficients, the $\hat{\beta}$s) in order to maximize its correlation with the response variable. Thus, a different composite would be used when the response variable is Mathematics than when it is Evidence-Based Reading and Writing, for example. Note that the $\hat{\alpha}_j$ term is not included in the equation. Again, this is because $\hat{y}_i$ represents prior achievement before the effect of the current district, school, or teacher.

The second step in the predictive model is to estimate the group effects (α_j) using the following ANCOVA model.

$$y_i = \gamma_0 + \gamma_1\hat{y}_i + \alpha_j + \epsilon_i \tag{19}$$

In the predictive model, the effects (α_j) are considered random effects. Consequently, the $\hat{\alpha}_j$s are obtained by shrinkage estimation (empirical Bayes).[4] The regression coefficients for the ANCOVA model are given by the γs.

2.3.3.1 Accommodations to the Predictive Model for Missing 2019-20 Data Due to the Pandemic

In spring 2020, the COVID-19 pandemic required schools to close early and cancel statewide summative assessments. As a result, statewide scores are not available for Michigan's M-STEP, PSAT and SAT exams based on the 2019-20 school year, and it is not possible to measure growth from the 2018-19 to the 2019-20 school years. For the predictive model, the 2020-21 reporting measures growth using students' predictors through the 2018-19 and then compares to their performance on the 2020-21 assessment.

As a reminder, the predictive model is used to measure growth for assessments given in non-consecutive grades, such as M-STEP Science and Social Studies assessments. Because these assessments are not administered every year, *it has always been possible* that students do not have any test scores in the *immediate* prior year. The model can provide a robust estimate of students' entering achievement for the course by using all other available test scores from other subjects, grades, and years.

In other words, the predictive model does not require any technical adaptations to account for the missing year of data and the interpretation of the results is similar to a typical year of reporting.

---

[4] For more information about shrinkage estimation, see, for example, Ramon C. Littell, George A. Milliken, Walter W. Stroup, Russell D. Wolfinger, and Oliver Schabenberger, *SAS for Mixed Models, Second Edition* (Cary, NC: SAS Institute Inc., 2006). Another example is Charles E. McCulloch, Shayle R. Searle, and John M. Neuhaus, *Generalized, Linear, and Mixed Models, Second Edition* (Hoboken, NJ: John Wiley & Sons, 2008).

2.4 Projection Model

2.4.1 Overview

The longitudinal data sets used to calculate growth measures for groups of students can also provide individual student projections to future assessments. A projection is reported as a probability of obtaining a specific score or above on an assessment, such as a 70% probability of scoring Proficient or above on the next summative assessment. The probabilities are based on the students' own prior testing history as well as how the cohort of students who just took the assessment performed. Due to the pandemic, the projections to the assessments for the 2021-22 school year are based on the cohort of students who took the assessment in the 2018-19 school year rather than the 2020-21 school year. Projections are available for state assessments as well as to college readiness assessments.

Projections are useful as a planning resource for educators, and they can inform decisions around enrollment, enrichment, remediation, counseling, and intervention to increase students' likelihood of future success.

2.4.2 Technical Description

The statistical model that is used as the basis for the projections is, in traditional terminology, an analysis of covariance (ANCOVA) model. This model is the same statistical model used in the predictive model applied at the school level described in Section 2.3.3. In the projection model, the score to be projected serves as the response variable (y), the covariates (x's) are scores on tests the student has already taken, and the categorical variable is the school at which the student received instruction in the subject, grade, and year of the response variable (y). Algebraically, the model can be represented as follows for the i^{th} student.

$$y_i = \mu_y + \alpha_j + \beta_1(x_{i1} - \mu_1) + \beta_2(x_{i2} - \mu_2) + \cdots + \epsilon_i \quad (20)$$

The μ terms are means for the response and the predictor variables. α_j is the school effect for the j^{th} school, the school attended by the i^{th} student. The β terms are regression coefficients. Projections to the future are made by using this equation with estimates for the unknown parameters (μs, βs, sometimes α_j). The parameter estimates (denoted with "hats," e.g., $\hat{\mu}$, $\hat{\beta}$) are obtained using the most current data for which response values are available. The resulting projection equation for the i^{th} student is

$$\hat{y}_i = \hat{\mu}_y \pm \hat{\alpha}_j + \hat{\beta}_1(x_{i1} - \hat{\mu}_1) + \hat{\beta}_2(x_{i2} - \hat{\mu}_2) + \cdots + \epsilon_i \quad (21)$$

The reason for the "±" before the $\hat{\alpha}_j$ term is that since the projection is to a future time, the school that the student will attend is unknown, so this term is usually omitted from the projections. This is equivalent to setting $\hat{\alpha}_j$ to zero, that is, to assuming that the student encounters the "average schooling experience" in the future.

Two difficulties must be addressed to implement the projections. First, not all students will have the same set of predictor variables due to missing test scores. Second, because this is an ANCOVA model with a school effect i, the regression coefficients must be "pooled-within-school" regression coefficients. The strategy for dealing with these difficulties is the same as described in Section 2.3.3 using equations (16), (17), and (18) and will not be repeated here.

Once the parameter estimates for the projection equation have been obtained, projections can be made for any student with any set of predictor values. However, to protect against bias due to measurement error in the predictors, projections are made only for students who have at least three available predictor scores (or, in the case of M-STEP Social Studies in grade 5 for the 2020-21 reporting, two predictor scores). In addition to the projected score itself, the standard error of the projection is calculated ($SE\,(\hat{y}_i)$). Given a projected score and its standard error, it is possible to calculate the probability that a student will reach some specified benchmark of interest (b). Examples are the probability of scoring at least Proficient on a future end-of-grade test or the probability of scoring at least an established college readiness benchmark. The probability is calculated as the area above the benchmark cutoff score using a normal distribution with its mean equal to the projected score and its standard deviation equal to the standard error of the projected score as described below. Φ represents the standard normal cumulative distribution function.

$$Prob(\hat{y}_i \geq b) = \Phi\left(\frac{\hat{y}_i - b}{SE\,(\hat{y}_i)}\right) \tag{22}$$

2.5 Outputs from the Models

2.5.1 Gain Model

The gain model is used for courses where students test in consecutive grade-given tests. As such, **the gain model uses M-STEP in Math and ELA in grades 3–7 and PSAT 8/9 in grade 8 to provide district, school, and teacher growth measures in the following content areas:**

- M-STEP Math in grades 5–7 for 2020-21 reporting, grades 4–7 for 2018-19 reporting, and grades 4–8 for 2017-18 reporting
- M-STEP ELA in grades 5–7 for 2020-21 reporting, grades 4–7 for 2018-19 reporting, and grades 4–8 for 2017-18 reporting
- PSAT 8/9 in grade 8 for 2018-19 and 2020-21 reporting

Note that teacher reporting is only available to those districts that have chosen to opt in through MiDataHub, and it can be based on either the statewide summative assessments listed above or the district's interim/benchmark assessments, which included the following in previous years:

- MAP Math in grades 1–8
- MAP Reading in grades 1–8

In addition to the mean scores and mean gain for an individual subject, grade, and year, the gain model can also provide the following teacher composites across subjects, grades, and years.

In general, these are all different forms of linear combinations of the random effects, and their estimates and standard errors are computed in the same manner described above in equations (9) and (10) for the teacher model. More details about teacher composites across subjects, grades, and years are available in Section 5.

2.5.2 Predictive Model

The predictive model is used for courses where students test in non-consecutive grade-given tests. As such, **the predictive model provides growth measures for districts and schools in the following content areas:**

- M-STEP Science in grades 5, 8 and 11
- M-STEP Social Studies in grades 5, 8 and 11
- PSAT 8/9 in grade 9
- PSAT 10
- SAT

In addition to the mean scores and growth measures for an individual subject, grade, and year, the predictive model can also provide multi-year average growth measures (up to three years) for each subject and grade or course.

2.5.3 Projection Model

Projections are provided to future state assessments as well as college readiness assessments:

- M-STEP Math and ELA in grades 5–7
- M-STEP Social Studies in grades 5, 8, and 11
- M-STEP Science in grades 8 and 11
- PSAT 8/9 Mathematics and ELA in grade 8
- PSAT 8/9 in Mathematics, Evidence-Based Reading and Writing in grade 9
- PSAT 10 Mathematics and Evidence-Based Reading and Writing in grade 10
- SAT Mathematics and Evidence-Based Reading and Writing

More specifically, M-STEP projections are typically provided one or two grade levels above a student's last tested grade, such as projections to grades 6 and 7 for students who most recently tested in grade 5. For the 2021 reporting, M-STEP projections are provided to a student's next tested grade-level M-STEP assessment, such as a projection to grade 6 for students who most recently tested in grade 5.

3 Expected Growth

3.1 Overview

Conceptually, growth is simply the difference between students' entering and exiting achievement. As noted in Section 2, zero represents "expected growth." Positive growth measures are evidence that students made *more* than the expected growth, and negative growth measures are evidence that students made *less* than the expected growth.

A more detailed explanation of expected growth and how it is calculated are useful for the interpretation and application of growth measures.

3.2 Technical Description

Both the gain and predictive models define expected growth based on the empirical student testing data; in other words, the model does not assume a particular amount of growth or assign expected growth in advance of the assessment being taken by students. Both models define expected growth within a year. This means that expected growth is always relative to how students' achievement has changed in the most recent year of testing rather than a fixed year in the past.

More specifically, **in the gain model, expected growth means that students maintained the same relative position with respect to the statewide student achievement that year. In the predictive model, expected growth means that students with a district, school, or teacher made the same amount of growth as students with the average district, school, or teacher in the state for that same year, subject, and grade**.

For both models, the growth measures tend to be centered on expected growth every year with approximately half of the district/school/teacher estimates above zero and approximately half of the district/school/teacher estimates below zero.

A change in assessments or scales from one year to the next does not present challenges to calculating expected growth. Through the use of NCEs, the gain model converts any scale to a relative position, and the predictive model already uses prior test scores from different scales to calculate the expected score. When assessments change over time, expected growth is still based on the relative change in achievement from one point in time to another.

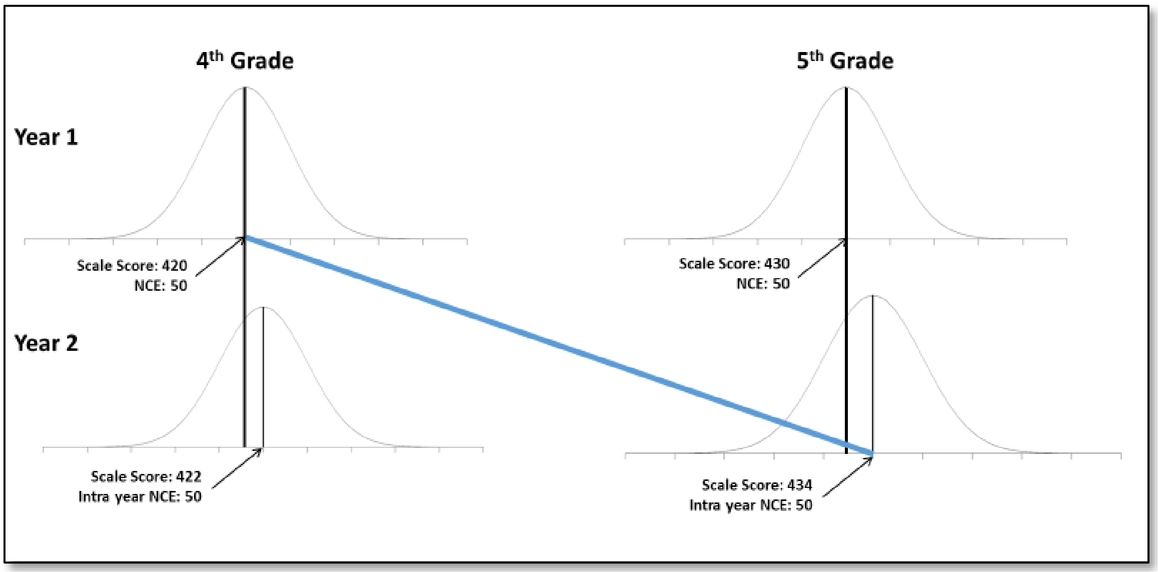
3.3 Illustrated Example

Figure 6 below provides a simplified example of how growth is calculated in the gain model when the state achievement increases. The figure has four graphs, each of which plots the NCE distribution of scale scores for a given year and grade. In this example, the figure shows how the gain is calculated for a group of grade 4 students in Year 1 as they become grade 5 students in Year 2. In Year 1, our grade 4 students score, on average, 420 scale score points on the test, which corresponds to the 50th NCE (similar to the 50th percentile). In Year 2, the students score, on average, 434 scale score points on the test, which corresponds to a 50th NCE *based on the grade 5 distribution of scores in Year 2*. The grade 5 distribution of scale scores in Year 2 was higher than the grade 5 distribution of scale scores in Year 1, which is why the lower right graph is shifted slightly to the right. The blue line shows what is required for students to make expected growth, which would be to maintain their position at the 50th NCE for grade

4 in Year 1 as they become grade 5 students in Year 2. The growth measure for these students is Year 2 NCE – Year 1 NCE, which would be 50 – 50 = 0. Similarly, if a group of students started at the 35th NCE, the expectation is that they would maintain that 35th NCE.

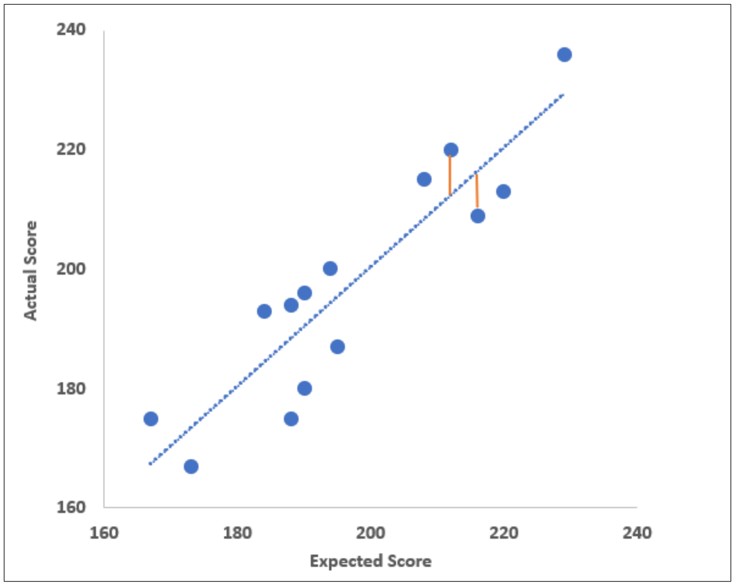
Note that the actual gain calculations are much more robust than what is presented here; as described in the previous section, the models can address students with missing data, team teaching, and all available testing history.

Figure 6: Intra-Year Approach Example for the Gain Model



In contrast, in the predictive model, expected growth uses actual results from the most recent year of assessment data and considers the relationships from the most recent year with prior assessment results. Figure 7 below provides a simplified example of how growth is calculated in the predictive model. The graph plots each student's actual score with their expected score. Each dot represents a student, and a best-fit line will minimize the difference between all students' actual and expected scores. Collectively, the best-fit line indicates what expected growth is for each student – given the student's expected score, expected growth is met if the student scores the corresponding point on the best-fit line. Conceptually, with the best-fit line minimizing the difference between all students' actual and expected scores, the growth expectation is defined by the average experience. Note that the actual calculations differ slightly since this is an ANCOVA model where the students are expected to see the average growth as seen by the experience with the average group (district, school, or teacher).

Figure 7: Intra-Year Approach Example for the Predictive Model



4 Classifying Growth into Categories

4.1 Overview

It can be helpful to classify growth into different levels for interpretation and context, particularly when the levels have statistical meaning. Michigan's growth model has five categories for districts and schools and four categories for teachers. These categories are defined by a range of values related to the growth measure, its standard error and (for Teacher reports) the student-level standard deviation of growth.

4.2 Use Standard Errors Derived from the Models

As described in the modeling approaches section, the growth model provides an estimate of growth for a district, school, or teacher in a particular subject, grade, and year as well as that estimate's standard error. The standard error is a measure of the quantity and quality of student data included in the estimate, such as the number of students and the occurrence of missing data for those students. It also takes into account shared instruction and team teaching. Standard error is a common statistical metric reported in many analyses and research studies because it yields important information for interpreting an estimate, in this case the growth measure relative to expected growth. Because measurement error is inherent in any growth or value-added model, *the standard error is a critical part of the reporting*. **Taken together, the growth measure and standard error provide educators and policymakers with critical information about the certainty that students in a district, school, or classroom are making decidedly more or less than the expected growth.** Taking the standard error into account is particularly important for reducing the risk of misclassification (for example, identifying a teacher as ineffective when they are truly effective) for high-stakes usage of value-added reporting.

The standard error also takes into account that even among teachers with the same number of students, teachers might have students with very different amounts of prior testing history. Due to this variation, the standard errors in a given subject, grade, and year could vary significantly among teachers, depending on the available data that is associated with their students, and it is another important protection for districts, schools, and teachers to incorporate standard errors to the value-added reporting.

4.3 Define Growth Indicators in Terms of Standard Errors

Common statistical usage of standard errors indicates the precision of an estimate and whether that estimate is statistically significantly different from an expected value. The growth reports use the standard error of each growth measure to determine the statistical evidence that the growth measure is different from expected growth. For EVAAS growth reporting, this is essentially when the growth measure is more than or less than two standard errors above or below expected growth or, in other words, when the growth index is more than +2 or less than -2. For district and school growth reports, these definitions then map to growth indicators in the reports themselves, such that there is statistical meaning in these categories. For the teacher growth reports, there is another statistic metric used to define growth categories, and this is described in the next section.

4.4 Define Growth Categories in Terms of Student-Level Standard Deviation of Growth

The student-level standard deviation of growth can be used to provide context about the magnitude of growth being made by a group of students. For the gain-based model (where this metric is applied), students typically have a current and a prior year NCE, which can be used to derive a student-level gain. The standard deviation of the student-level distribution of growth is available for each year, subject, and grade. Dividing the growth measures by the standard deviation provides a value known as an "effect size," and it indicates the practical significance regarding the group of students and whether they met, exceeded, or fell short of expected growth.

The categories and definitions for district, school, and teacher growth reports are illustrated in the following section.

4.5 Categorizing District and School Growth Measures

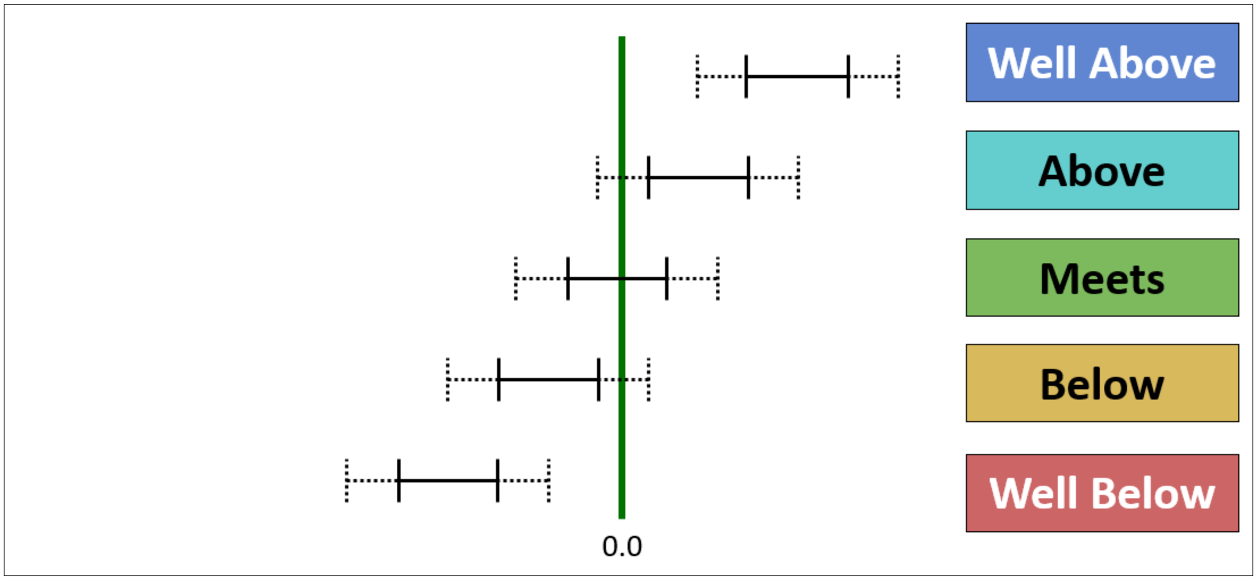
There are two ways to visualize how the growth measure and standard error relate to expected growth and how these can be used to create categories.

The first way is to frame the growth measure relative to its standard error and expected growth at the same time. For district and school reporting, the categories are defined as follows:

- **Well Above** indicates that the growth measure is two standard errors or more above expected growth (0). This level of certainty is significant evidence that students made more growth than expected.
- **Above** indicates that the growth measure is at least one but less than two standard errors above expected growth (0). This is moderate evidence that students made more growth than expected.
- **Meets** indicates that the growth measure is less than one standard error above expected growth (0) but no more than two standard errors below expected growth (0). This is evidence that students made growth as expected.
- **Below** indicates that the growth measure is more than one but no more than two standard errors below expected growth (0). This is moderate evidence that students made less growth than expected.
- **Well Below** is an indication that the growth measure is less than or equal to two standard errors below expected growth (0). This level of certainty is significant evidence that students made less growth than expected.

Figure 8 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 Well Above 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 Well Below category. Above and Below 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. Meets indicates that there is evidence that students made growth as expected as both the solid and dotted cross the line indicating expected growth.

Figure 8: Visualization of Growth Categories with Expected Growth, Growth Measures, and Standard Errors



This graphic is helpful in understanding how the growth measure relates to expected growth and whether the growth measure represents a statistically significant difference from expected growth.

The second way to frame the categories is to create a growth index, which is calculated as shown below:

$$Growth\ Index = \frac{Growth\ Measure - Expected\ Growth}{Standard\ Error\ of\ the\ Growth\ Measure} \tag{23}$$

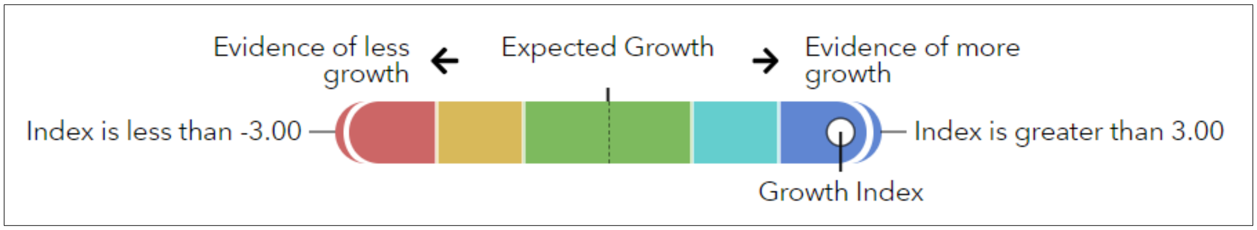
The growth index is similar in concept to a Z-score or t-value, and it communicates as a single metric the certainty or evidence that the growth measure is decidedly above or below expected growth. The growth index is useful when comparing value-added measures from different assessments or in different units, such as NCEs or scale scores. The categories can be established as ranges based on the growth index, such as the following:

- **Well Above** indicates significant evidence that students made more growth than expected. The growth index is 2 or greater.
- **Above** indicates moderate evidence that students made more growth than expected. The growth index is between 1 and 2.
- **Meets** indicates evidence that students made growth as expected. The growth index is between -1 and 1.
- **Below** indicates moderate evidence that students made less growth than expected. The growth index is between -2 and -1.

- **Well Below** indicates significant evidence that students made more growth than expected. The growth index is less than -2.

This is represented in the growth indicator bar in Figure 9, which is similar to what is provided in the District and School Value-Added reports in the EVAAS web application. The black dotted line represents expected growth. The color-coding within the bar indicates the range of values for the growth index within each category.

Figure 9: Sample Growth Indicator Bar



It is important to note that these two illustrations provide users with the same information; they are simply presenting the growth measure, its standard error, and expected growth in different ways.

4.6 Categorizing Teacher Growth Measures

Teacher reporting will categorize teacher growth measures using a two-step process based on, first, the growth index and, second, the effect size.

Again, the growth index is the growth estimate divided by the standard error, which is specific to each estimate. The effect size is the growth measure divided by the student-level standard deviation of growth. The effect size provides an indicator of magnitude and practical significance that the group of students met, exceeded, or fell short of expected growth.

This two-step approach first considers whether there is statistical certainty that the growth measure is above or below the expectation of growth. The second step determines whether the growth measure is above or below the growth expectation by a certain magnitude. The first step uses the growth index to determine thresholds for the certainty, and the second step uses the effect size to determine thresholds for magnitude.

For the first step with uncertainty, the thresholds are an index of +2 or greater, an index of -2 or less, or an index between -2 and +2. These thresholds are similar to the concept of a 95% confidence interval. If a 95% confidence interval around the growth measure did not contain the growth expectation, then they would fall outside the thresholds. The second step uses an effect size threshold of 0.4 and -0.4. These values correspond to a "medium" effect size as referenced in John Hattie's work.[5]

In accordance with MDE policies, there are four categories for teacher growth categorization. The top category has a growth index of greater than or equal to 2 *and* an effect size of greater than or equal to 0.4. The next highest category consists of all other measures where the growth index is greater than or

[5] See, for example, John Hattie, *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement* (London: Routledge, 2008).

equal to -2 *and* one other condition is met: either the index is also less than 2 *or* the effect size is less than 0.4. The bottom category is when the growth index is less than -2 and the effect size is less than -0.4. The next to bottom category are teachers with a growth index less than -2, but their effect size is greater than or equal to -0.4.

The table below provides the color-coding, definitions, and interpretation for the Value-Added reports of *teachers*.

Table 3: Teacher Value-Added Categories, Definitions, and Interpretations

Category	Definition	Interpretation
Level 4, Exceeds	Index is greater than or equal to 2 *and* the effect size is greater than or equal to 0.40	Level 4, Exceeds: Significant evidence that the teacher's students made more progress than the growth standard and the effect size is medium or higher
Level 3, Met	Index is greater than or equal to -2 *and* either the index is less than 2 or the effect size is less than 0.4.	Level 3, Met: Evidence that the teacher's students made progress similar to the growth standard
Level 2, Nearly Met	Index is less than -2 *and* the effect size is greater than or equal to -0.4.	Level 2, Nearly Met: Significant evidence that the teacher's students made less progress than the growth standard but not less than a negative medium effect size
Level 1, Not Met	Index is less than -2 *and* the effect size is less than -0.4.	Level 1, Not Met: Significant evidence that the teacher’s students made less progress than the growth standard and less growth than a negative medium effect size

NOTE: When an index falls exactly on the boundary between two colors, the higher growth color is assigned.

4.7 Rounding and Truncating Rules

As described in the previous section, the effectiveness level is based on the value of the growth index. As additional clarification, the calculation of the growth index uses unrounded values for the value-added measures and standard errors. After the growth index has been created but before the categories are determined, the index values are rounded or truncated by taking the maximum value of the rounded or truncated index value out to two decimal places. This provides the highest category given any type of rounding or truncating situation. For example, if the score was a 1.995, then rounding would provide a higher category. If the score was a -2.005, then truncating would provide a higher category. In practical terms, this impacts only a very small number of measures.

Also, when value-added measures are combined to form composites, as described in the next section, the rounding or truncating occurs after the final index is calculated for that combined measure.

5 Composite Growth Measures

Updated teacher reports have not been released yet, so this section describes composites available through the 2019-20 school year.

A composite combines growth measures from different subjects, grades, and/or years for an individual teacher. Teacher reporting is available for the M-STEP Math and ELA assessments and PSAT 8/9 in grade 8 as well as MAP Math and Reading in grades 1–8. It is not available for the other PSAT and SAT assessments because they are not course-specific and tend to assess general content knowledge that would be covered in several courses at the high school level. Teacher reporting is available for the 2017-18 and 2018-19 school years for state summative assessments and for the 2017-18, 2018-19, and 2019-20 (MOY) school years for MAP.

Teachers will receive a composite if they have teacher reporting available for the most recent year of reporting (2018-19 for state summative assessments and 2019-2020 [MOY] for MAP assessments). Depending on what is available for the teacher, the following composites are available:

- Subject-specific composite across grades for a given type of test, such as:
  - Up to three-year M-STEP Math for grades 4–7 and PSAT 8/9 in grade 8
  - Up to three-year M-STEP ELA for grades 4–7 and PSAT 8/9 in grade 8
  - Up to three-year MAP Math for grades 1–8
  - Up to three-year MAP Reading for grades 1–8
- Overall composite across subjects and grades for a given type of test, such as:
  - Up to three-year M-STEP Math and ELA for grades 4–7 and PSAT 8/9 in grade 8
  - Up to three-year MAP Math and Reading for grades 1–8

Note that these composites are based on one type of assessment, state summative *or* MAP, not both combined. Based on MDE policy, these composites will include up-to-three years of growth data. If a teacher only has one year of growth data for the most recent year, then that teacher's composite only includes growth data from that single year.

The key policy decisions for combining growth measures can be summarized as follows:

- A composite is weighted by the number of "full-time equivalent" students associated with each individual growth measure for the type of assessment (state summative or MAP).
- For each teacher, the full-time equivalent (FTE) number of students is based on the number of students linked to that teacher as well as the percentage of instructional time the teacher has for each student. For example, if a teacher taught 26 students for 50% of their instructional time, then the teacher's student FTE number would be 26 students times 50% of their instructional learning time, or 13 students.
- Typically, this growth is combined within a year first and then across years.
- The across-year measures are also weighted by the student FTE number.

The following sections show how a composite is calculated for a sample teacher.

5.1 Teacher Composites

5.1.1 Overview

The key steps for determining a teacher's composite index are as follows:

1. Calculate the *gain* across grades and subjects for a given year.
2. Calculate the *standard error* across grades and subjects for a given year.
3. Calculate the composite *gain* across years.
4. Calculate the composite *standard error* across years.
5. Calculate the composite *index* across years.
6. Calculate the composite *effect size* across years.

If a teacher does not have multiple years of value-added measures, then the composite index would be based on the single-year composite index.

The following sections illustrate this process using value-added measures from a sample teacher, which are provided in Table 4.

Table 4: Sample Teacher Value-Added Information

Year	Subject	Grade	Growth Measure	Standard Error	Index	Std. Dev.	Effect Size	Number of FTE Students
2019	Math	6	3.30	0.70	4.71	11.0	0.30	25
2019	ELA	6	-1.10	1.00	-1.10	10.0	-0.11	23
2021	Math	6	1.70	0.65	2.62	10.5	0.16	27

5.1.2 Technical Description of the Composite Index Based on Gain Model Measures

The composite index for the gain model growth measures is calculated by dividing the composite gain by its composite standard error. The calculations for each of these metrics are provided below.

5.1.2.1 Calculate Gain Across Grades and Subjects for a Given Year

Because all growth measures from the gain-based model are in the same scale (Normal Curve Equivalents), the teacher composite gain across the two applicable subject/grades is a weighted average of the individual gains based on the number of effective students in each subject and grade. For the teacher, the total number of FTE students affiliated with gain-based growth measures in 2019 is 25 + 23, or 48. The 2019 grade 6 Math value-added measure would be weighted at 25/48, the 2019 grade 6 ELA value-added measure would be weighted at 23/48. More specifically, the composite gain is calculated using the following formula:

$$2019\ Comp\ Gain = \frac{25}{48} Math_6 + \frac{23}{48} ELA_6 = \frac{25}{48}(3.30) + \frac{23}{48}(-1.10) = 1.19 \tag{20}$$

5.1.2.2 Calculate Standard Error Across Grades and Subjects for a Given Year

5.1.2.2.1 Technical Background on Standard Errors

The standard error of the gain-based teacher composite gain cannot be calculated using the assumption that the gains making up the composite are independent. This is because many of the same students are likely represented in different value-added gains, such as grade 6 Math in 2019 and grade 6 ELA in 2019. The statistical approach, outlined in Section 3.1.3 (with references), is quite sophisticated and will consider the correlations between pairs of value-added gains as shown in equation (21) below and using equation (10) for teachers.[6] The composites are indeed linear combinations of the fixed effects of the models and can be estimated as described in Section 3.1.3. The magnitude of each correlation depends on the extent to which the same students are in both estimates for any two subject, grade, and year estimates.

5.1.2.2.2 Illustration of Gain-Based Standard Error for Sample Teacher

As a reminder, the use of the word "error" does not indicate a mistake. Rather, growth/value-added models produce *estimates*. The growth measures in the above tables are estimates of the teacher's true value-added effectiveness based on student test score data. In statistical terminology, a "standard error" is a measure of the uncertainty in the estimate, providing a means to determine whether an estimate is decidedly above or below the growth expectation. Standard errors can, and should, also be provided for the composite gains that have been calculated.

Statistical formulas are often more conveniently expressed as variances, and this is the square of the standard error. Standard errors of composites can be calculated using variations of the general formula shown below. To maintain the generality of the formula, the individual estimates in the formula (think of them as value-added gains) are simply called X, Y, and Z. If there were more than or fewer than three estimates, the formula would change accordingly. As gain-based composites use proportional weighting according to the number of FTE students linked to each value-added gain, each estimate is multiplied by a different weight: a, b, or c.

$$\begin{aligned} Var(aX + bY + cZ) &= a^2 Var(X) + b^2 Var(Y) + c^2 Var(Z) \\ &+2ab\, Cov(X,Y) + 2ac\, Cov(X,Z) + 2bc\, Cov(Y,Z) \end{aligned} \tag{21}$$

Covariance, denoted by Cov, is a measure of the relationship between two variables. It is a function of a more familiar measure of relationship, the correlation coefficient. Specifically, the term $Cov(X,Y)$ is calculated as follows:

$$Cov(X,Y) = Correlation(X,Y)\sqrt{Var(X)}\sqrt{Var(Y)} \tag{22}$$

The value of the correlation ranges from -1 to +1, and these values have the following meanings:

- A value of zero indicates no relationship.
- A positive value indicates a positive relationship, or Y tends to be larger when X is larger.

[6] For more details about the statistical approach to derive the standard errors, see, for example, Ramon C. Littell, George A. Milliken, Walter W. Stroup, Russell D. Wolfinger, and Oliver Schabenberger, *SAS for Mixed Models, Second Edition* (Cary, NC: SAS Institute Inc., 2006). Another example: Charles E. McCulloch, Shayle R. Searle, and John M. Neuhaus, *Generalized, Linear, and Mixed Models* (Hoboken, NJ: Wiley, 2008).

- A negative value indicates a negative relationship, or Y tends to be smaller when X is larger.

Two variables that are unrelated have a correlation and covariance of zero. Such variables are said to be statistically independent. If the X and Y values have a positive relationship, then the covariance will also be positive. As a general rule, two value-added gain estimates are statistically independent if they are based on completely different sets of students.

For our sample teacher's composite gain, the relationship will generally be positive, and this means that the gain-based composite standard error is larger than it would be assuming independence. Using the student weightings and standard errors reported in Table 4 and assuming total independence, the standard error would then be as follows:

$$\begin{aligned} 2018\ Comp\,SE &= \sqrt{\left(\frac{25}{48}\right)^2 (SE\ Math_6)^2 + \left(\frac{23}{48}\right)^2 (SE\ ELA_6)^2} \\ &= \sqrt{\left(\frac{25}{48}\right)^2 (0.70)^2 + \left(\frac{23}{48}\right)^2 (1.00)^2} = 0.60 \end{aligned} \tag{23}$$

At the other extreme, if the correlation between each pair of value-added gains had its maximum value of +1, the standard error would be larger.

In this example, since the teacher teaches the same grade in different subjects, the actual standard error will likely be above the value of 0.60 due to students being in both Math and ELA with the teacher. The specific value will depend on the values of the correlations across the two gains. Correlations of gains across years might be positive or slightly negative since the same student's score can be used in multiple gains if a teacher has taught that student multiple times. The magnitude of each correlation depends on the extent to which the same students are in both estimates for any two subject/grade/year estimates.

For the sake of simplicity, let us assume the actual standard error was 0.65 for the teacher composite in this example.

5.1.2.3 Calculate Composite Gain Across Years

The next step is to calculate the gain for students across time for this teacher. The composite gain would be found by taking the weighted average of year's gain as follows:

$$Comp\ gain = \frac{48}{75} gain_{2019} + \frac{27}{75} gain_{2021} = \frac{48}{75}(1.19) + \frac{27}{75}(1.70) = 1.37 \tag{24}$$

Although some of the values in the example were rounded for display purposes, the actual rounding or truncating only occurs after all of measures have been combined, as described in Section 4.7.

5.1.2.4 Calculate Composite Standard Error Across Years

The calculations above provide the composite gain across years. Then we have a standard error for each year. These can be combined to create a standard error for the composite gain. Assuming independence across time and using the student weightings and single-year standard errors, the multi-year standard error would then be as follows:

$$Comp\ SE = \sqrt{\left(\frac{48}{75}\right)^2 (SE_{2019})^2 + \left(\frac{27}{75}\right)^2 (SE_{2021})^2} = \sqrt{\left(\frac{48}{75}\right)^2 (0.60)^2 + \left(\frac{27}{75}\right)^2 (0.65)^2} = 0.45$$

5.1.2.5 Calculate Composite Index Across Years

The next step is to calculate the teacher composite index, which is the teacher composite value-added gain divided by its standard error. The gain-based composite index for this teacher would be calculated as follows:

$$Comp\ Index = \frac{Comp\ Gain}{Comp\ SE} = \frac{1.37}{0.45} = 3.04 \qquad (24)$$

Although some of the values in the example were rounded for display purposes, the actual rounding or truncating only occurs after all of measures have been combined as described in Section 4.7.

5.1.2.6 Calculate the Composite Effect Size Across Years

To calculate the effect size for the overall composite, each growth measure is divided by the student-level standard deviation of growth. This value is a constant within each year subject and grade but can be different across the different year, subject, and grades. The composite effect size is a weighted average of the effect sizes based on the FTE number of students.

$$\begin{aligned} Comp\ Effect\ Size &= \frac{25}{75} Math_{2019_6} + \frac{23}{75} ELA_{2019_6} + \frac{27}{75} Math_{2021_6} \\ &= \frac{25}{75}(0.30) + \frac{23}{75}(-0.11) + \frac{27}{75}(0.16) = 0.12 \end{aligned} \qquad (25)$$

5.1.2.7 Categorizing Growth Measures as a Final Step

With the combined composite growth index and effect size, the specific composite can be categorized. This growth index is above 2.00. The effect size is below 0.40. Therefore, based on Section 4.6, the teacher composite would fall into Level 3 or Met.

6 Input Data Used in the Michigan Growth Model

6.1 Assessment Data Used in Michigan

For the analysis and reporting based on the 2020-21 school year, EVAAS received the following assessments for use in the growth and/or projection models:

- M-STEP English Language Arts (ELA) and Mathematics in grades 3–7
- M-STEP Science in grades 5, 8, and 11
- M-STEP Social Studies in grade 5, 8, and 11
- PSAT 8/9 in Mathematics and ELA in grade 8
- PSAT 8/9 in Mathematics, Evidence-Based Reading and Writing in grade 9
- PSAT 10 in Mathematics, Evidence-Based Reading and Writing in grade 10
- SAT in Mathematics, Evidence-Based Reading and Writing in grade 11

These assessments are administered in the spring semester of the school year.

EVAAS received interim/benchmark assessments from districts that opted to submit them for teacher value-added reporting through MiDataHub, and the following assessments met the criteria for assessments in Section 7.1 as well as minimum number requirements in Section 7.3.2:

- MAP Mathematics in grades 1–8
- MAP Reading in grades 1–8

These assessments are administered each year at the beginning of year (BOY), middle of year (MOY) and end of year (EOY). BOY includes test scores from August through October, MOY includes test scores from December through February, and EOY includes test scores from March through June. Note that, for the 2019-20 school year, only BOY and MOY are available.

State assessment files from MDE provided the following data for each student score:

- Scale score
- Test taken
- Tested grade
- Tested subject
- Tested semester
- Tested performance level
- Full Academic Year designation
- Educational Entity Master District code
- Educational Entity Master District name
- Educational Entity Master School code
- Educational Entity Master School name

Some of this information, such as performance levels, is not relevant to PSAT or SAT tests.

Interim/benchmark assessment files from the MiDataHub provided the following endpoints and required data elements:

- academicSubjectDescriptors
    - codeValue

    - description/shortDescription
    - namespace
- assessments
    - assessmentIdentifier
    - namespace
    - assessmentFamily
    - assessmentForm
    - assessmentTitle
    - assessmentVersion
    - academicSubjectDescriptor
    - gradeLevelDescriptor
- gradeLevelDescriptors
    - codeValue
    - description/shortDescription
    - namespace
- schools
    - schoolId
    - nameOfInstitution
    - operationalStatusDescriptor
    - schoolTypeDescriptor
    - shortNameOfInstitution
    - localEducationAgencyId
    - educationOrganizationIdentificationSystemDescriptor
    - identificationCode
- studentAssessments
    - studentAssessmentIdentifier
    - administrationDate
    - administrationEndDate
    - administrationLanguageDescriptor
    - whenAssessedGradeLevelDescriptor
    - assessmentIdentifier
    - namespace
    - schoolYear
    - studentUniqueId
    - performanceLevels.assessmentReportingMethodDescriptor
    - performanceLevelDescriptor
    - performanceLevelMet
    - scoreResults.assessmentReportingMethodDescriptor
    - scoreResults.resultDatatypeTypeDescriptor
    - scoreResults
- students
    - studentUniqueId
    - birthDate
    - firstName
    - lastSurname
    - middleName

- studentSchoolAssociations
  - entryDate
  - entryGradeLevelDescriptor
  - exitWithdrawDate
  - schoolReference.schoolId
  - schoolYearTypeReference.schoolYear
  - studentUniqueId

More information about these endpoints is available in the Ed-Fi Operational Data Store API.

6.2 Student Information

Student information is used in creating the web application to assist educators analyze the data to inform practice and assist all students with academic growth. SAS receives this information in the form of various socioeconomic, demographic, and programmatic identifiers provided by MDE. SAS received the following student information and identifiers from MDE:

- Gender (Male, Female, Unknown)
- Race
  - American Indian or Alaska Native
  - Asian
  - Black or African American
  - Hispanic or Latino
  - Native Hawaiian or Other Pacific Islander
  - Two or More Races
  - Unknown
  - White
- Economically Disadvantaged (Y, N) – only reported at aggregate levels
- English Learner (Y, N)
- Special Education (Y, N)
- Homeless (Y, N) – only reported at aggregate levels

6.3 Teacher Information

It is possible for Michigan educators to receive Teacher growth reports from EVAAS. To provide these reports, SAS must receive teacher information from the MiDataHub to use in conjunction with MDE's student assessment data and the local interim/benchmark assessment data. This is necessary since the EVAAS models estimate the teacher growth measures for the group of students that are connected to a teacher in a given subject and grade. To receive this information, districts, and/or ISDs must opt in to share MiDataHub data with SAS.

6.3.1 Data Used for Teacher-Student Linkages

The MiDataHub project contains different data tables for various purposes. EVAAS uses the following tables to obtain data for teacher-student linkages and/or identify the name and code of the district:

- schools
- students
- studentSectionAssociations

- staffSectionAssociations
- staffs
- courses
- courseOfferings
- academicSubjectDescriptors
- classroomPositionDescriptors
- CalendarDates
- Schools
- LocalEducationAgencies

The last three tables are only used to validate the district during the opt-in process.

Within these endpoints, SAS uses the following elements:

- academicSubjectDescriptors
  - codeValue
  - namespace
  - description/shortDescription
- classroomPositionDescriptors
  - codeValue
  - namespace
  - shortDescription
  - courseOfferings
  - localCourseCode
  - localCourseTitle
  - courseCode
  - schoolId
  - schoolYear
  - sessionName
  - courses
  - courseCode
  - courseTitle
  - academicSubjectDescriptor
  - schools
  - schoolId
  - nameOfInstitution
  - operationalStatusDescriptor
  - schoolTypeDescriptor
  - shortNameOfInstitution
  - localEducationAgencyId
  - educationOrganizationIdentificationSystemDescriptor
  - identificationCode
  - staffs
  - staffUniqueId
  - firstName
  - lastSurname
  - electronicMailAddress
  - electronicMailTypeDescriptor

- staffIdentificationSystemDescriptor
- identificationCode
- staffSectionAssociations
- beginDate
- endDate
- classroomPositionDescriptor
- teacherStudentDataLinkExclusion
- localCourseCode
- schoolId
- schoolYear
- sectionIdentifier
- sessionName
- staffUniqueId
- students
- studentUniqueId
- birthDate
- firstName
- lastSurname
- middleName
- studentSectionAssociations
- beginDate
- endDate
- teacherStudentDataLinkExclusion
- localCourseCode
- schoolId
- schoolYear
- sectionIdentifier
- sessionName
- studentUniqueId

6.3.2 Assigning Subject Areas

EVAAS uses the "localCourseTitle" element from the "courseOfferings" and "courses" endpoints to categorize courses as either Math or ELA for the state assessments. EVAAS also references the available subject area descriptions via the "AcademicSubjectDescriptors" endpoint and retains all records with subject areas relevant to our assessment pool (ELA/MATH).

Below are some examples of the values EVAAS looks for to identify the subject area that a course falls into. This list is not exhaustive.

- Mathematics: MATH, MTH, ALGEBRA, ALG, GEOMETRY
- English Language Arts: English, ELA, LANGUAGE ARTS, LANG ARTS, READ, LIT, L ARTS, LA

If EVAAS was not able to find teacher data for assessments for an entire grade in a school, the course names that were received for these students were reviewed. In these cases, course names such as "Homeroom" or "GRADE 4" are categorized as both Math and ELA since these were most likely self-contained classrooms.

If courses were not categorized as either Math or ELA, corresponding records were dropped. Course names that were not categorized into either subject area were excluded. For example, course names that only referenced "Spelling," "Grammar," or "Writing" were not included.

6.3.3 Calculating Instructional Responsibility

EVAAS uses the student and teacher start and end dates to calculate how much of a student's instruction in a subject each teacher that interacted with that student is responsible for.

Percentages of instructional responsibility are based on two things:

- The number of days a teacher taught a student in a tested subject compared to the total number of days the student was enrolled in the subject.
- The number of teachers who taught the student.

Capturing the proportion of instructional responsibility for each teacher at the individual student level ensures EVAAS Teacher Value-Added reports link student growth to teachers fairly and accurately.

When calculating instructional responsibility, EVAAS uses the start and end dates for a student and teacher to determine how long a teacher provided instruction to a student in a course/subject. A student appears on a teacher's roster if the student's start and end dates overlap with the teachers.

If a student appears on multiple teachers' rosters for the same subject at the same time, instructional responsibility is split across the teachers. Here are two scenarios:

- Bobby is in a year-long grade 5 Math course, and Mrs. Smith is the only teacher of record for that course for all the days that Bobby is in that class. Mrs. Smith's instructional responsibility is 100%.
- Two teachers co-teach Bobby's year-long grade 5 Math course for the entire year. Each teacher has 50% instructional responsibility for Bobby.

In addition, EVAAS calculates the proportion of the school year each student received instruction in the tested subject. If a student was not enrolled for the entire school year, EVAAS adjusts the teacher's instructional responsibility to reflect the student's shortened instruction time. For example:

- Bobby's family moves to the area, and he enrolls in Mrs. Smith's year-long grade 5 Math course on the 45th day of the 180-day school calendar. Because he was in the class for 135 days and Mrs. Smith has 100% instructional responsibility, Mrs. Smith has 75% of the instructional responsibility for him.
- Bobby's family moves to the area, and he enrolls in Mrs. Smith and Mr. Jones' split year-long grade 5 Math course on the 45th day of the 180-day school calendar. Because he was in the class for 135 days and he splits his time with Mrs. Smith and Mr. Jones, each teacher has 37.5% of the instructional responsibility for him.
- Mrs. Smith was hired to replace Mr. Jones on the 90th day of the 180-day school calendar. Mrs. Smith and Mr. Jones have 50% instructional responsibility for the class.

Other business rules that affect the linkage data include:

- If a teacher's start and end dates are not populated, EVAAS uses the start and end dates for the student that the teacher is connected to.
- Courses that fall under the umbrella of Math or ELA are linked to the corresponding students' test scores. However, if a student is linked to both a general grade-level Math teacher and a

Geometry teacher, EVAAS only links that student to the general grade-level Math teacher. If no general Math teacher exists, EVAAS links the student to the Geometry teacher.

- For courses that fall under the umbrella of Math or ELA, EVAAS attributes students' enrollment to their assessments. This means that the instructional responsibility for a student is split across multiple teachers if a student is enrolled in multiple ELA courses simultaneously.
- EVAAS excludes courses that are not discernable as Math or ELA from analysis. There are exceptions to this exclusion rule for districts, schools, and grades that have very low linkage rates or in cases where the courses table can be used to manually verify the subject area correctly indicates ELA or Mathematics.

6.3.4 Records Dropped in Initial Processing for Teacher-Student Linkages

There are several reasons why student and teacher data submitted through the tables in the MiDataHub might be removed through EVAAS' initial data processing. Some examples are listed below.

- No live data within MiDataHub at the time of the pull for the required endpoints.
- EVAAS connects data from the "students" endpoint and the "studentSectionAssociations" endpoint using UIC (studentUniqueId). If a UIC is present in one endpoint and not the other, then the record is incomplete and will be excluded.
- The course information provided in the "studentSectionAssociations" endpoint must have connecting course information present in the "staffSectionAssociations" endpoint so that a teacher record and student record can be connected. If the course information is present in one endpoint but not the other, then those records are excluded.
- If EVAAS is unable to identify these course titles referencing the "courseOfferings" or "courses" endpoint, then these courses are dropped. This connection does not exist, and the records are excluded. The course information present in our "studentSectionAssociations" and "staffSectionAssociations" endpoints must have identifiable course name information within the "courseOfferings" or "courses" endpoint. If EVAAS is unable to identify these courses with titles referencing the "courseOfferings" or "courses" endpoint, then these courses are dropped.
- If a value of studentUniqueID does not exist in any assessment data that EVAAS has received, then those records are excluded.
- If teacher and student dates do not overlap, then those records are removed from processing.

6.3.5 Combining Teacher-Student Linkages with Assessment Records

Once there is final set of teacher-student linkages, that information is connected to the assessment records to be used in the teacher value-added models. Students will have to meet other requirements described in the remainder of this document to be included in the teacher's growth measure.

7 Business Rules

7.1 Assessment Verification for Use in Growth Models

To be used appropriately in any growth models, the scales of these assessments must meet three criteria:

1. **There is sufficient stretch in the scales** to ensure growth can be measured for both low-achieving students as well as high-achieving students. A floor or ceiling in the scales could disadvantage educators serving either low-achieving or high-achieving students.
2. **The test is highly related to the academic standards** so that it is possible to measure growth with the assessment in that subject, grade, and year.
3. **The scales are sufficiently reliable from one year to the next.** This criterion typically is met when there are a sufficient number of items per subject, grade, and year. This will be monitored each subsequent year that the test is given.

These criteria are checked annually for each assessment prior to use in any growth model, and Michigan's current standardized assessments meet them. These criteria are explained in more detail below.

7.1.1 Stretch

Stretch indicates whether the scaling of the assessment permits student growth to be measured for both very low- or very high-achieving students. A test "ceiling" or "floor" inhibits the ability to assess students' growth for students who would have otherwise scored higher or lower than the test allowed. It is also important that there are enough test scores at the high or low end of achievement, so that measurable differences can be observed.

Stretch can be determined by the percentage of students who score near the minimum or the maximum level for each assessment. If a much larger percentage of students scored at the maximum in one grade than in the prior grade, then it might seem that these students had negative growth at the very top of the scale when it is likely due to the artificial ceiling of the assessment. Percentages for all Michigan assessments are well below acceptable values, meaning that these assessments have adequate stretch to measure value-added even in situations where the group of students are very high or low achieving.

7.1.2 Relevance

Relevance indicates whether the test is sufficiently aligned with the curriculum. The requirement that tested material correlates with standards will be met if the assessments are designed to assess what students are expected to know and be able to do at each grade level. More information about Michigan academic standards can be found at the following link: https://www.michigan.gov/mde/0,4615,7-140-28753---,00.html

7.1.3 Reliability

Reliability can be viewed in a few different ways for assessments. Psychometricians view reliability as the idea that a student would receive similar scores if the assessment was taken multiple times. The type of reliability is important for most any use of standardized assessments. This criterion typically is met when there is a sufficient number of items per subject/grade/year, and this will be monitored each subsequent year that the test is given.

7.2 Pre-Analytic Processing

7.2.1 Missing Grade

In Michigan, the grade used in the analyses and reporting is the tested grade, not the enrolled grade. If a grade is missing on any M-STEP or MAP assessments, then that record will be excluded from all analyses. The grade is required to include a student's score in the appropriate part of the models and to convert the student's score into the appropriate NCE in the gain-based model.

7.2.2 Duplicate (Same) Scores

If a student has a duplicate score for a particular subject and tested grade in a given testing period in a given school, then the extra score will be excluded from the analysis and reporting.

7.2.3 Students with Missing Districts or Schools for Some Scores but Not Others

If a student has a score with a missing district or school for a particular subject and grade in a given testing period, then the duplicate score that has a district and/or school will be included over the score that has the missing data.

7.2.4 Students with Multiple (Different) Scores in the Same Testing Administration

If a student has multiple scores in the same period for a particular subject and grade and the test scores are not the same, then those scores will be excluded from the analysis. For MAP assessments, if a student has multiple scores in the same period, the BOY test score is defined as the first test date for a student/test/subject/grade and the MOY and EOY test scores are defined as the last test date for a student/test/subject/grade. If a student has multiple scores in the same period for a particular subject and grade and the test scores are not the same, then those scores will be excluded from the analysis. This is applied to state assessments and any remaining MAP assessment records that could not be resolved by the first and last test date business rule.

If duplicate scores for a particular subject and tested grade in a given testing period are at different schools, then both scores will be excluded from the analysis.

7.2.5 Students with Multiple Grade Levels in the Same Subject in the Same Year

A student should not have different tested grade levels in the same subject in the same year. If that is the case, then the student's records are checked to see whether the data for two separate students were inadvertently combined. If this is the case, then the student data are adjusted so that each unique student is associated with only the appropriate scores. If the scores appear to all be associated with a single unique student, then scores that appear inconsistent are excluded from the analysis.

7.2.6 Students with Records That Have Unexpected Grade Level Changes

If a student skips more than one grade level (e.g., moves from sixth in year 1 to ninth in year 2) or is moved back by one grade or more (i.e. moves from fourth in year 1 to third in year 2) in the same subject, then the student's records are examined to determine whether two separate students were inadvertently combined. If this is the case, then the student data is adjusted so that each unique student is associated with only the appropriate scores. These scores are removed from the analysis if it is the same student.

7.2.7 Students with Records at Multiple Schools in the Same Test Period

If a student is tested at two different schools in a given testing period, then the student's records are examined to determine whether two separate students were inadvertently combined. If this is the case, then the student data is adjusted so that each unique student is associated with only the appropriate scores. When students have valid scores at multiple schools in different subjects, all valid scores are used at the appropriate school.

7.2.8 Outliers

Student assessment scores are checked each year to determine whether they are outliers in context with all the other scores in a reference group of scores from the individual student. These reference scores are weighted differently depending on proximity in time to the score in question. Scores are checked for outliers using related subjects as the reference group. For example, when searching for outliers for Math test scores on state assessments, all Math scores from state assessments are examined simultaneously during outlier identification for the state assessments, and any scores that appear inconsistent, given the other scores for the student, are flagged. Outlier identification for college readiness assessments use all available college readiness data alongside state assessments in the respective subject area (e.g., Math subjects with M-STEP and PSAT tests might be used to identify outliers with SAT).

Scores are flagged in a conservative way to avoid excluding any student scores that should not be excluded. Scores can be flagged as either high or low outliers. Once an outlier is discovered, that outlier will not be used in the analysis, but it will be displayed on the student testing history on the EVAAS web application.

This process is part of a data quality procedure to ensure that no scores are used if they were, in fact, errors in the data, and the approach for flagging a student score as an outlier is fairly conservative.

Considerations included in outlier detection are:

- Is the score in the tails of the distribution of scores? Is the score very high or low achieving?
- Is the score "significantly different" from the other scores as indicated by a statistical analysis that compares each score to the other scores?
- Is the score also "practically different" from the other scores? Statistical significance can sometimes be associated with numerical differences that are too small to be meaningful.
- Are there enough scores to make a meaningful decision?

To decide whether student scores are considered outliers, all student scores are first converted into a standardized normal Z-score. Then each individual score is compared to the weighted combination of all the reference scores described above. The difference of these two scores will provide a t-value of each comparison. Using this t-value, the growth models can flag individual scores as outliers.

There are different business rules for the low outliers and the high outliers, and this approach is more conservative when removing a very high-achieving score.

For low-end outliers, the rules are:

- The percentile of the score must be below 50.

- The t-value must be below -3.5 for M-STEP grades 3–7 and PSAT 8/9 in grade 8 for Math and ELA and for MAP grades 1–8 for Math and Reading when determining the difference between the score in question and the weighted combination of reference scores (otherwise known as the comparison score). In other words, the score in question must be at least 3.5 standard deviations below the comparison score. For other assessments, the t-value must be below -4.0
- The percentile of the comparison score must be above a certain value. This value depends on the position of the individual score in question but will range from 10 to 90 with the ranges of the individual percentile score.

For high-end outliers, the rules are:

- The percentile of the score must be above 50.
- The t-value must be above 4.5 for M-STEP grades 3–7 and PSAT 8/9 in grade 8 for Math and ELA and for MAP grades 1–8 for Math and Reading when determining the difference between the score in question and the reference group of scores. In other words, the score in question must be at least 4.5 standard deviations above the comparison score. For other assessments, the t-value must be above 5.0.
- The percentile of the comparison score must be below a certain value. This value depends on the position of the individual score in question but will need to be at least 30 to 50 percentiles below the individual percentile score.
- There must be at least three scores in the comparison score average.

7.2.9 Linking Records over Time

Each year, EVAAS receives data files that include student assessment data and file formats. These data are checked each year prior to incorporation into a longitudinal database that links students over time. Student test data and demographic data are checked for consistency year to year to ensure that the appropriate data are assigned to each student. Student records are matched over time using all data provided by the state, and teacher records are matched over time using the Unique ID and teacher's name.

7.3 Growth Models

7.3.1 Students Included in the Analysis

As described in Section 7.2 (Pre-Analytic Processing), student scores might be excluded due to the business rules, such as outlier scores.

For the gain model, all students are included in these analyses if they have assessment scores that can be used. The gain model uses all available M-STEP and PSAT 8/9 in grade 8 for Math and ELA results for each student for M-STEP growth measures and all available MAP Math and Reading results for MAP growth measures. For the M-STEP growth measures, student scores are excluded if they are flagged to indicate that they did not meet Full Academic Year. MAP scores are excluded if there is not a BOY and an EOY score for the reporting year (or a BOY and an MOY score for 2019-20 reporting).

Because this model follows students from one grade to the next and measures growth as the change in achievement from one grade to the next, the gain model assumes typical grade patterns for students. Students with non-traditional patterns, such as those who have been retained in a grade or skipped a grade, are treated as separate students in the model. In other words, these students are still included in the gain model, but the students are treated as separate students in different cohorts when these non-traditional patterns occur. This process occurs separately by subject since some students can be accelerated in one subject and not in another.

For the predictive and projection models, a student must have at least three valid predictor scores that can be used in the analysis, all of which cannot be deemed outliers. (See Section 7.2.8 on Outliers.) These scores can be from any year, subject, and grade that are used in the analysis. In other words, the student's expected score can incorporate other subjects beyond the subject of the assessment being used to measure growth. The required three predictor scores are needed to sufficiently dampen the error of measurement in the tests to provide a reliable measure. If a student does not meet the three-score minimum, then that student is excluded from the analyses. It is important to note that not all students have to have the same three prior test scores; they only have to have some subset of three that were used in the analysis. Unlike the gain model, students with non-traditional grade patterns are included in the predictive model as one student. Since the predictive model does not determine growth based on consecutive grade movement on tests, students do not need to stay in one cohort from one year to the next. That said, if a student is retained and retakes the same test, then that prior score on the same test will not be used as a predictor for the same test as a response in the predictive model. This is mainly due to the fact that very few students used in the models have a prior score on the same test that could be used as a predictor. In fact, in the predictive model, it is typically the case that a prior test is only considered a possible predictor when at least 50% of the students used in that model have those prior test scores. Student scores are excluded from the predictive model if they are flagged to indicate that they did not meet Full Academic Year. There are no membership rules used to include or exclude students in the PSAT 8/9 in grade 9, PSAT 10, SAT and MAP analyses.

7.3.2 Minimum Number of Students to Receive a Report

The growth models require a minimum number of students in the analysis in order for districts, schools, and teachers to receive a growth report. This is to ensure reliable results.

7.3.2.1 District and School Model

For the gain model, the minimum student count to report an estimated average NCE *score* (i.e., either entering or exiting achievement) is seven students in a specific subject, grade, and year. To report an estimated NCE *gain* in a specific subject, grade, and year, there are additional requirements:

- Of those students who are associated with the school or district in the current year and grade, there must be at least seven students in each subject, grade, and year in order for that subject, grade, and year to be used in the gain calculation.
- There is at least one student at the school or district who has a "simple gain," which is based on a valid test score in the current year and grade as well as the prior year and grade in the same subject. However, due to the rule above, it is typically the case that at least seven students have a "simple gain." In some cases where students only have a Math or Reading score in the current year or previous year, this value dips below seven.
- For any district or school growth measures based on specific student groups, the same requirements described above apply for the students in that specific student group.

For example, to report an estimated NCE gain for school A in M-STEP Math grade 5 for this year, there must be the following requirements:

- There must be at least seven fifth-grade students with an M-STEP Math grade 5 score at school A for this year.
- Of the fifth-grade students at school A this year *in all subjects, not just Math*, there must be at least seven students with an M-STEP Math grade 4 score from last year.
- At least one of the fifth-grade students at school A this year must have an M-STEP Math grade 5 score from this year *and* an M-STEP Math grade 4 score from last year.

For the predictive model, the minimum student count to receive a growth measure is seven students in a specific subject, grade, and year. These students must have the required three prior test scores needed to receive an expected score in that subject, grade, and year.

7.3.2.2 Teacher Model

The teacher gain *model* includes teachers who are linked to at least seven students with a valid test score in the same subject, grade, and year. This requirement does not consider the percentage of instructional time that the teacher spends with each student in a specific subject and grade.

To receive a Teacher *report* in a particular year, subject, and grade, there is an additional requirement. A teacher must have at least five Full Time Equivalent (FTE) students in a specific subject, grade, and year for the state assessments or in a specific subject, grade, and semester for the MAP assessment. The teacher's number of FTE students is based on the number of students linked to that teacher and the percentage of instructional time the teacher has for each student. For example, if a teacher taught 10 students for 50% of their instructional time, then the teacher's FTE number of students would be five, and the teacher would not receive a teacher growth report. If another teacher taught 14 students for 50% of their instructional time, then that teacher would have seven FTE students and would receive a teacher growth report. The instructional time attribution is obtained from the student-teacher linkage data described in Section 6.3.

The teacher gain model has an additional requirement. The teacher must be linked to at least seven students with prior test score data in the same subject, and the test data can come from any prior grade as long as they are part of the student's regular cohort. One of these seven students must have a "gain," meaning the same subject prior test score must come from the immediate prior year and prior grade for state assessments or the beginning of year semester of the current year and grade for MAP assessments. Students are linked to a teacher based on the subject area taught and the assessment taken. Students that have no prior testing data in the same subject area are not linked to the teacher for the analysis. Note that if a student repeats a grade, then the prior test data would not apply as the student has started a new cohort.

7.4 Student-Teacher Linkages

Student-teacher linkages are connected to assessment data based on the subject and identification information described in Section 6.3. The model will make adjustments to linkages if a student is claimed by teachers at a total percentage higher than 100% in an individual year, subject, and grade. If over-claiming happens, then the individual teacher's weight is divided by the total sum of all weights to redistribute the attribution of the student's test scores across teachers. Underclaimed linkages for students are not adjusted because a student can be claimed less than 100% for various reasons (such as a student who lives out of state for part of the year).