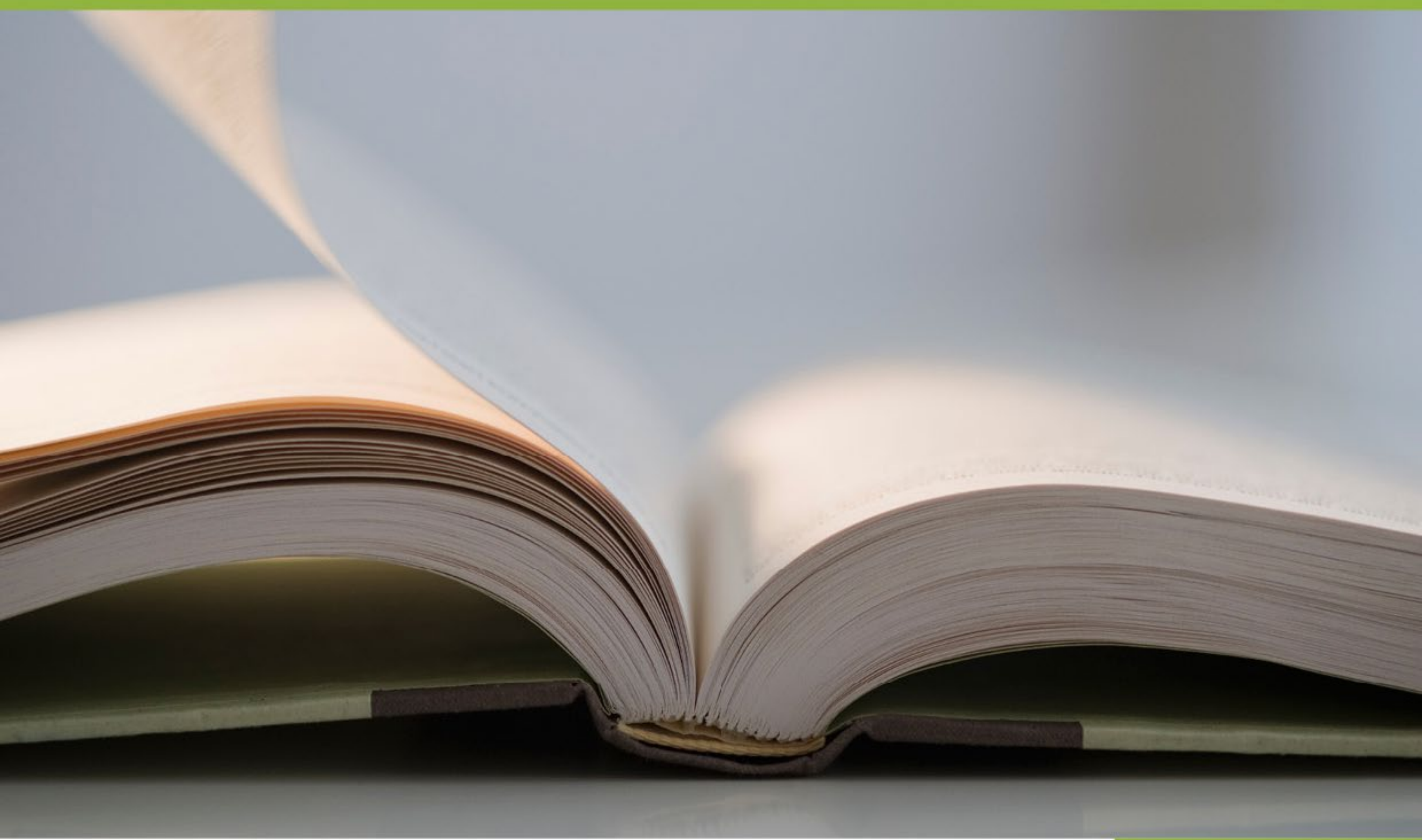
SAS® EVAASTopics in PVAAS Reporting for Pennsylvania



Contents

Introduction

Since 2002, SAS has provided Pennsylvania educators and policymakers with a powerful tool to determine—grade by grade and subject by subject—whether all students have plentiful choices and increased opportunities for learning. Known as PVAAS, these analyses follow the progress of individual students over time to:

- Assess a group of students' growth in LEAs/districts, schools, and classrooms; and
- Provide trajectories for individual students toward critical academic benchmarks.

Through the Pennsylvania Department of Education (PDE), this reporting is available to every public LEA/district and school based on the statewide Pennsylvania System of School Assessments (PSSA) and Keystones as well as a range of other national, state, and local assessments, by means of a secure web application.

The value-added estimates and student-level projections provided by PVAAS are based on robust and reliable methodologies. This important approach overcomes many critical statistical issues related to using standardized tests to assess student growth and student-level indicators as well as mitigating concerns about fairness. However, there are often questions about the details and implications of PVAAS value-added estimates. This document seeks to address those topics by describing details of the PVAAS methodology in theory and how they impact reporting in practice.

The information in this document is based on the typical PVAAS reporting and includes results from the 2018-19 school year. Due to the pandemic's impact on student learning, the models, interpretation, and results from the 2020-21 and 2021-22 reporting might differ somewhat than what is described here.

Topics Related to the Student Population Served by Educators

Topic: It is harder to make growth with students from certain demographic or socioeconomic backgrounds.

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

However, this adjustment is not statistically necessary for the most sophisticated value-added models, such as those used for PVAAS in the Commonwealth of Pennsylvania. This is because PVAAS uses all available testing history for each individual student and does not exclude students who have missing test data. Each student serves as their own control, and to the extent that SES/DEM influences persist over time, these influences are already represented in the student's data.

PVAAS in Theory

As a 2004 Ed Trust study stated, specifically with regard to the SAS EVAAS modeling, which is the approach used in Pennsylvania's LEA/district, school, and teacher reporting:

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

This finding has been confirmed independently by prominent value-added experts who have replicated a variety of value-added models, including PVAAS models. For example, in a 2007 paper by RAND researchers J.R. Lockwood and Dan McCaffrey explicitly verified the PVAAS modeling approach and described them as "extremely effective" at reducing bias in estimates of teacher's contributions to growth.[2] UCLA researchers Kilchan Choi, Pete Goldschmidt, and Kyo Yamashiro provided a similar finding in their study comparing value-added models:

> First, adding in an adjustment for student SES (as measured by eligibility for free- or reduced-price lunch) adds very little once a student's initial status is controlled… This indicates that student initial status captures many of the effects that SES is attempting to measure. In other words, by controlling for initial status, the model already captures the preceding effects that SES might have on students.[3]

In essence, these independent researchers have found that a sophisticated value-added approach does not typically systematically advantage or disadvantage educators by the type of students that they serve. By including so many prior test scores for each student, the model controls for many student characteristics that might impact their entering achievement or growth throughout the year.

---

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

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

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

Ultimately, there might be additional political and policy considerations that lead policymakers to make socioeconomic or demographic adjustments in the value-added models, but sophisticated ones tend neither to advantage nor disadvantage educators regardless.

PVAAS in Practice

Although the statistical literature presents evidence that educators are not advantaged or disadvantaged by the type of students that they serve in sophisticated value-added reporting, actual data offers more meaningful evidence to support this belief. The figures below provide teacher-level data, and the results are similar to those for LEA/district- and school-level.

The graph in Figure 1 plots the percentage of tested students who are considered economically disadvantaged for a specific teacher's roster in Pennsylvania against a teacher's growth index (the value-added estimate divided by its standard error) for PSSA Mathematics in grade 7 in 2023. Each dot represents one teacher, and verified rosters were used where available. Regardless of the student characteristics served by teacher, there is little to no correlation to the growth index. In other words, the dots representing each teacher do not trend up or down as the percentage increases; the cluster of dots is fairly even across the spectrum. In the graph below, the actual correlation between the growth index and percentage of economically disadvantaged students is -0.001, which is negligible.

Figure 1: Pennsylvania Growth Index Versus Percent Tested Economically Disadvantaged by Teacher

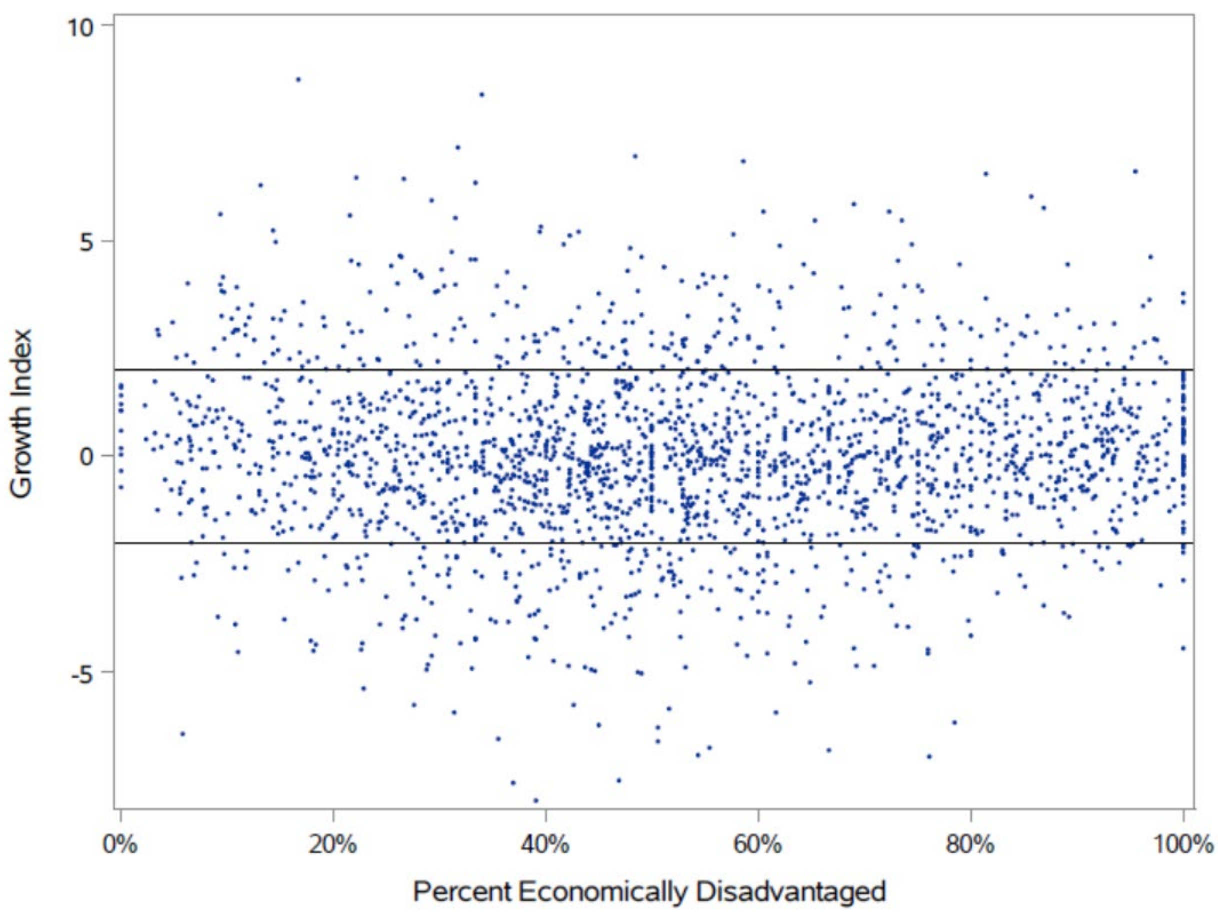


Figure 2 provides similar information for the percentage of students considered English Learner (EL), and there is little to no correlation to the growth index. In the graph below, the actual correlation between the growth index and percentage of students testing as EL is 0.07, which is negligible.

Figure 2: Pennsylvania Growth Index Versus Percent Tested EL by Teacher

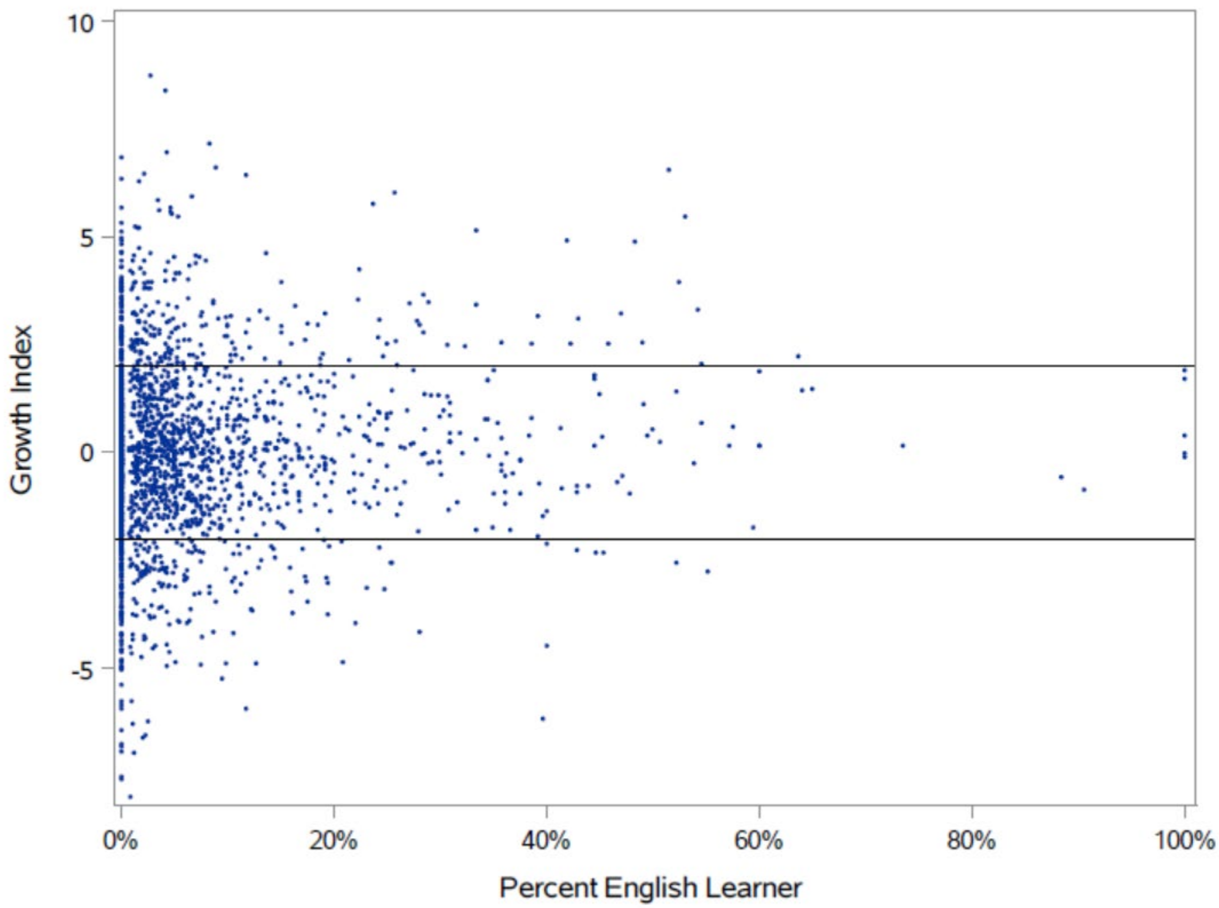
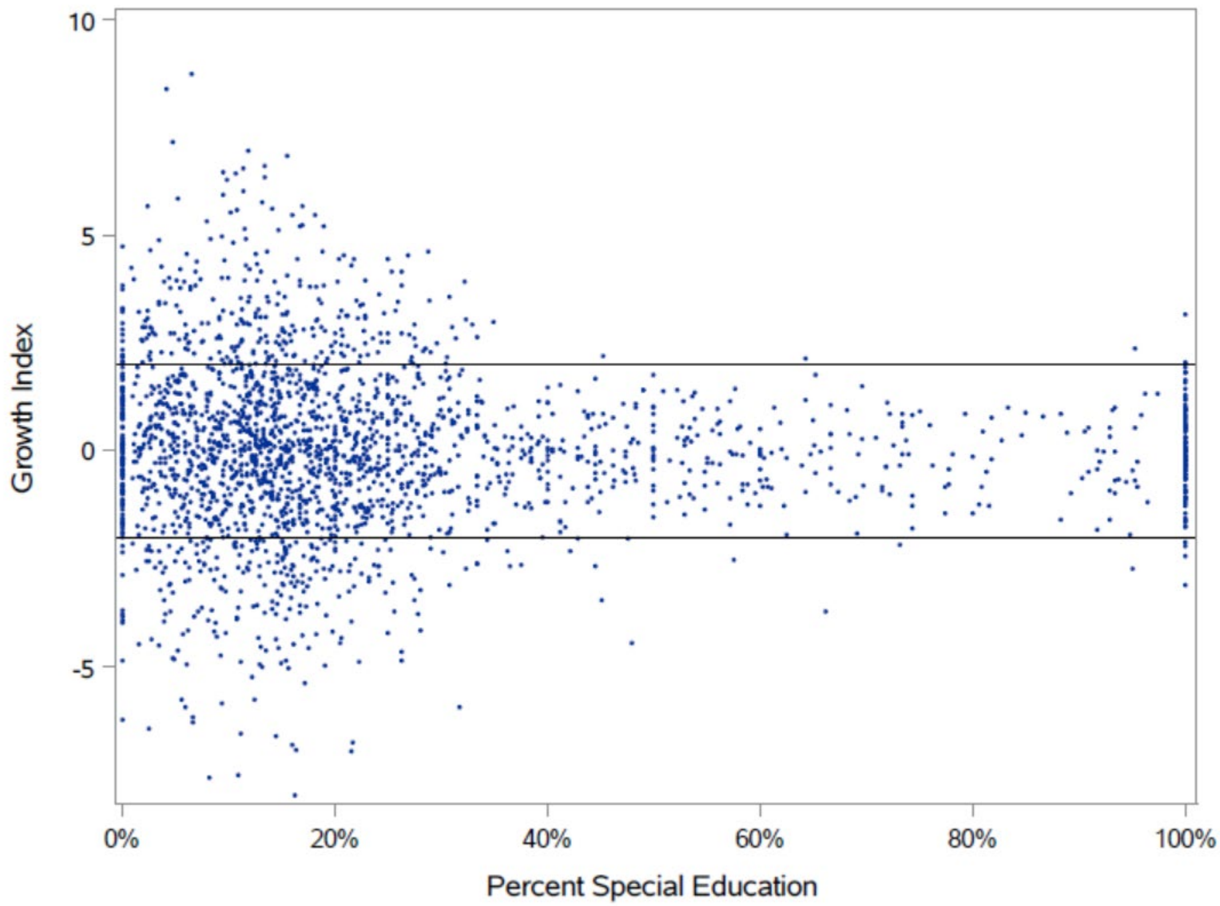


Figure 3 provides similar information for the percentage of students considered special education, and there is little to no correlation to the growth index. In the graph below, the actual correlation between the growth index and percentage of students testing as special education is -0.04, which is negligible.

Figure 3: Pennsylvania Growth Index Versus Percent Tested Special Education by Teacher



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

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

PVAAS in Theory

The value-added models used in Pennsylvania are designed to estimate whether a group of students made enough progress to meet the growth standard, which is based on a comparison of the group's average achievement to their average prior achievement.

Although Pennsylvania state assessments are designed to discriminate proficient from non-proficient, the state assessments are also designed to have sufficient stretch to discriminate between Below Basic, Basic, Proficient and Advanced performance levels. Accordingly, there is sufficient stretch in the state testing scales to differentiate performance and measure the growth of students with histories of higher achievement and those with histories of lower achievement.

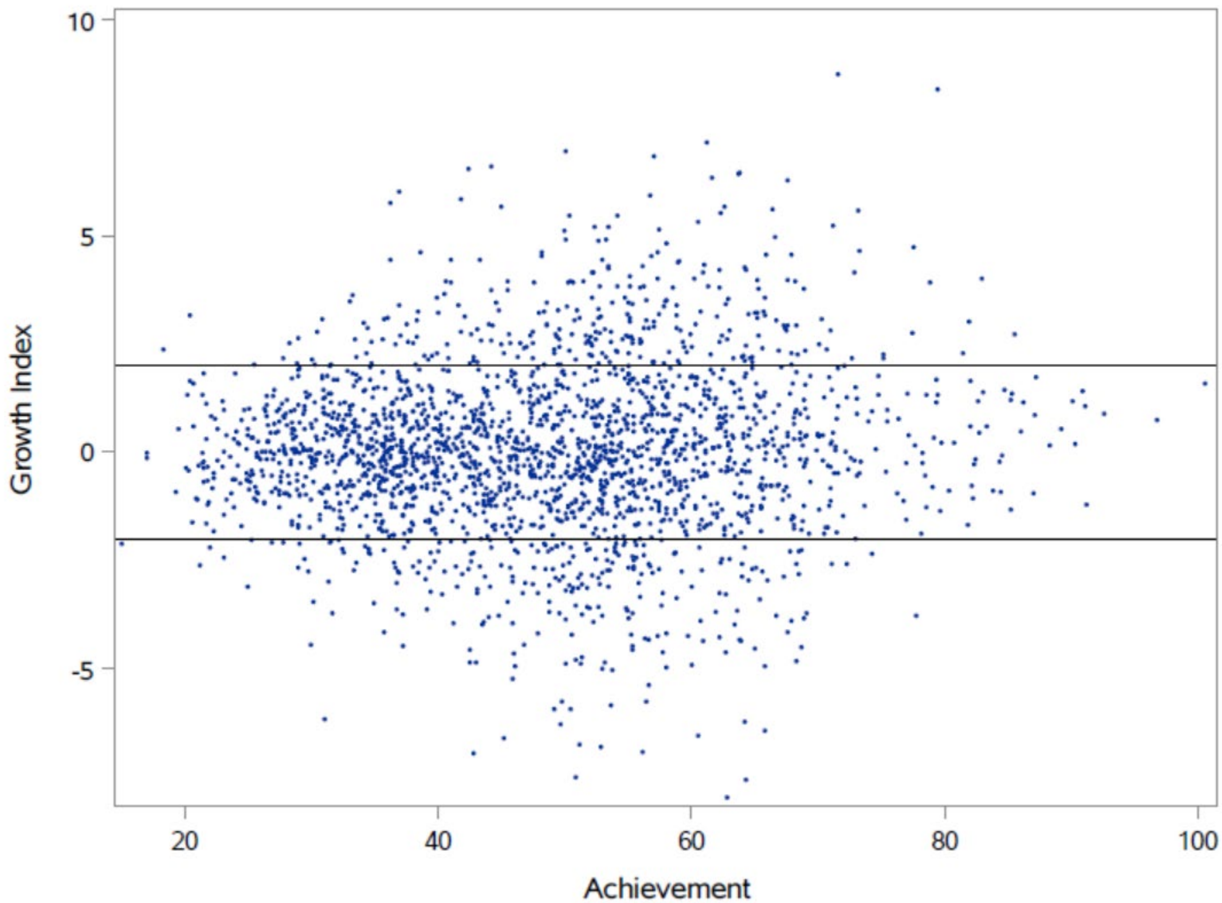
It is a requirement that any test that is used in PVAAS analyses meet the criteria of demonstrating sufficient stretch at the extremes. This requirement ensures that progress can be measured for both low-achieving students as well as high-achieving students.

PVAAS is fair not only to LEAs/districts, schools, and teachers serving students with a history of higher achievement, but it is also fair to the students themselves. The modeling that underlies PVAAS considers the growth of all students, regardless of their entering achievement, and the reporting shows whether the curriculum and instruction is targeted appropriately to these students. Students with a history of higher achievement might require enrichment work and more rigor in the same way that students with a history of lower achievement might need remediation to make sufficient growth.

PVAAS in Practice

Actual data can be used as evidence that growth measures are not influenced by the achievement level of the students served by LEA/districts, schools, and teachers. The graph in Figure 4 plots the average achievement for the students served by an individual teacher in Pennsylvania against a growth index (the value-added estimate divided by its standard error) for PSSA Mathematics in grade 7 in 2023. Each dot represents one teacher. Regardless of the teacher's student achievement, there is little to no correlation to the growth index. In other words, the dots representing each teacher do not trend up or down as achievement increases; the cluster of dots is fairly even across the achievement spectrum. In the following graph, the actual correlation between the growth index and entering achievement is 0.07, which is weak. LEA/District and school value-added plots are similar to the teacher one shown below. This shows that students with a history of higher achievement can—and do—show growth through PVAAS, and that educators are neither advantaged nor disadvantaged by the achievement level of their students.

Figure 4: Pennsylvania Growth Index Versus Average Achievement by Teacher



Topic: PVAAS should always indicate growth if the percentage of students scoring Proficient or above increased since last year.

Academic proficiency is in an important metric for measuring student success and increasing the number of proficient students is a goal of many educational entities. However, measuring only the percentage of students who score Proficient (or above) over time does not account for changes in achievement within performance level categories. PVAAS value-added reporting follows the progress of individual students (as a part of a group of students) over time, regardless of their achievement level, to ensure that all students count.

PVAAS in Theory

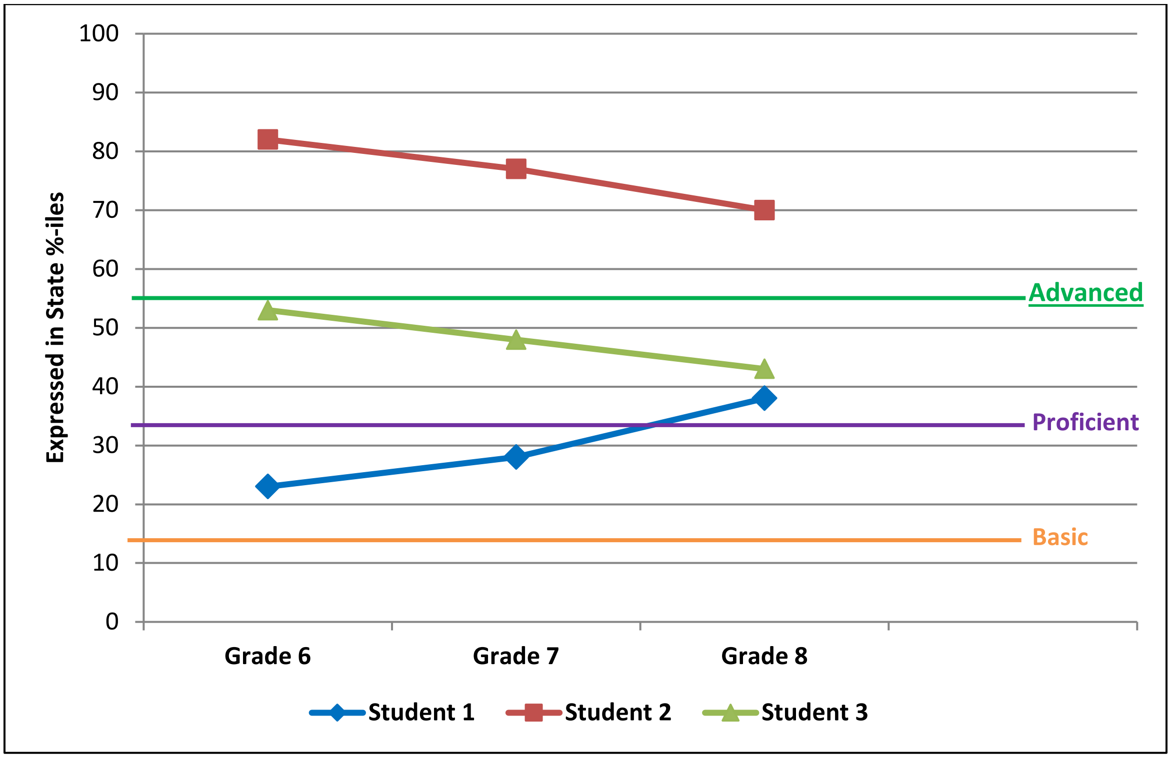
To demonstrate the limitations of focusing on the percentage of students who are proficient, consider the mathematics achievement history of three students illustrated in Figure 5 below:

- Student 1 is represented by the line with the blue diamonds. This student's achievement has steadily increased over time, with the student moving from Basic to Proficient from seventh to eighth grade.
- Student 2 is represented by the line with the red squares. This student's achievement has been steadily declining although the student has remained in the Advanced performance category.

- Student 3's achievement is represented by the line with the green triangles, and their achievement has been steadily declining but has stayed within the Proficient threshold.

The orange, purple, and green lines show the percentile corresponding to the Basic, Proficient, and Advanced performance levels.

Figure 5: Student Testing History in PSSA Mathematics for Student 1, Student 2, and Student 3



By considering only the change in proportion of students who met or exceeded the Proficient threshold, it would appear that this group of three students is generally improving because the number of students has increased with the addition of Student 1. However, this does not consider that the achievement level of the other two students is steadily decreasing over time. Alternatively, looking at growth would indicate that, as a group, students are now performing worse than expected. This example helps illustrate why a subtler approach is required that would consider the growth of all students, regardless of their achievement level.

Note: PVAAS does not provide a measure of growth for individual students, only for groups of students.

PVAAS in Practice

PVAAS does not measure students' growth based on the number or percentage of students who tested Below Basic, Basic, Proficient, or Advanced as compared to previous years. PVAAS detects changes in growth both across and within performance levels. As a result, educators are recognized when they make growth with students at/above proficiency and below proficiency. This can be very encouraging to LEAs/districts, schools, and teachers serving students with a history of lower achievement who might not otherwise be recognized for their students' growth.

Topic: PVAAS cannot measure the growth of students with missing data or highly mobile students.

Some measures of student growth cannot account for students with missing data. However, PVAAS value-added analyses provide reliable and valid estimates of students' growth with LEAs/districts, schools, and teachers, including those with high mobility. This is because PVAAS can include students even if they have missing test data, so that the growth is representative of the students actually served by LEAs/districts, schools, and teachers.

PVAAS in Theory

Including students with missing test scores as part of growth estimates is important for accurate and meaningful growth estimates because, in practice, highly mobile students are more likely to have a history of lower achievement. Excluding these students from the analysis could provide misleading growth estimates to LEAs/districts, schools, and teachers. More simplistic value-added or growth estimates might require that all students have the exact same set of prior test scores or that students have all prior test scores. However, this often has the result of excluding mobile student populations, and this would disproportionately affect educators serving those types of students.

To counteract this, PVAAS does not require that students have the same set of prior scores, which means PVAAS can include more students in calculating the growth measures. When estimating students' entering achievement, the modeling considers the quantity and quality of information available to each student as well as student mobility among schools from year to year.

To accomplish this without imputing student test scores, PVAAS uses a sophisticated modeling approach that provides more reliable estimates of growth.[4] The approach used by PVAAS for PSSA Mathematics and ELA estimates the means in each of these cells using relationships between students' test scores as if there were no missing test scores. In this way, the model provides more reliable and less biased growth measures without imputing any data. Furthermore, PVAAS uses much more student data to obtain these relationships in the growth estimates for LEAs/districts, schools, and teachers.

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

PVAAS in Practice

For PSSA Mathematics 3–8 and ELA 3–8, all students can be included regardless of their testing history, their number of prior test scores, and which test scores they have. For PSSA Science grade 8 and the Keystones, all students can be included as long as they have three prior test scores in any test, grade, and subject. For grade 4 Science, students need two prior test scores in any test, grade, and subject.

At PDE's request, student-level business rules are applied including not meeting the full academic year (FAY) requirement.

Because PVAAS reporting is available statewide in Pennsylvania, students and their testing histories can be followed as they move within the Commonwealth.

[4] 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://pvaas.sas.com/support/EVAAS-AdvantagesOfAMultivariateLongitudinalApproach.pdf

Topics Related to the Tests Used in Value-added Modeling

Topic: PVAAS reporting is not reliable or valid since it is based only on the PA state assessments.

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

PVAAS in Theory

Student test scores are the basic ingredient of all PVAAS analyses. EVAAS is not involved in and has no control over test construction. Pennsylvania's assessment system performs a universal assessment of Pennsylvania standards, and the assessments are aligned to the appropriate standards that are sufficient for longitudinal modeling and prediction. Regardless, before using any tests in PVAAS modeling, rigorous data processing and analyses verify that the tests meet the following three criteria:

- They demonstrate sufficient stretch so that students with a history of lower achievement and those with a history of higher achievement can show growth.
- They are aligned to state curriculum standards.
- The scales are sufficiently reliable from year to year.

To date, Pennsylvania's state assessments have met these criteria. More specifically, EVAAS analyses verify that there are enough different scale scores at the top and bottom of the scales to differentiate student achievement. EVAAS processing also analyzes the percentage of students scoring at the top and bottom scores to ensure there are no ceilings or floors. After all analyses are completed and PVAAS growth measures are available, SAS verifies that LEAs/districts, schools, and teachers serving students with histories of both higher and lower achievement can show both high and low growth. This process is repeated every year. More details about how the models underlying PVAAS support measuring growth among high and low achieving students can be found on pages 5-6.

Another common concern of educators is that they might be held accountable for how students did on a single test on a given day. EVAAS understands this concern and agrees that any single score represents only a snapshot of student performance at a particular point in time. However, the use of many test scores across subjects, grades, and years in PVAAS can provide a more complete picture of student learning and how students' achievement has changed over time.

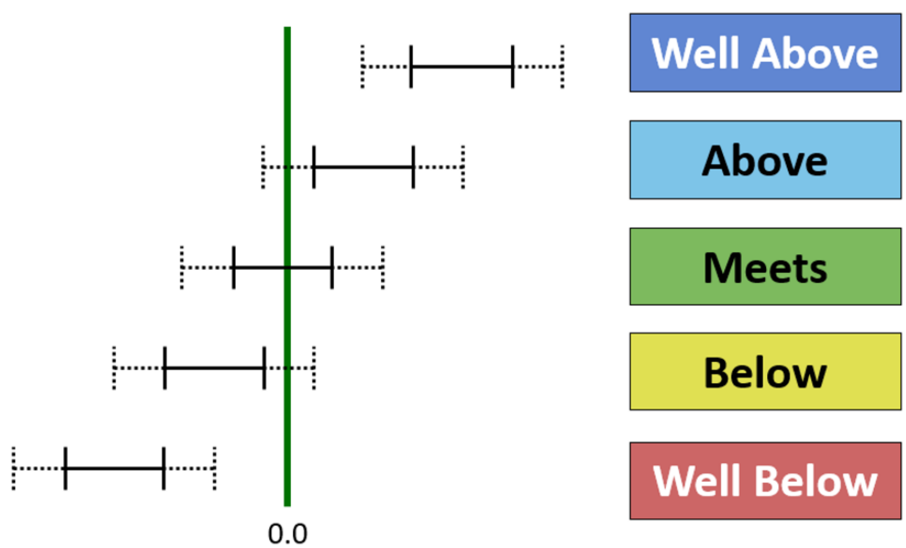
More specifically, the modeling approaches that underly PVAAS reporting are designed to minimize the impacts of the inherent measurement error associated with standardized tests on estimates of growth. First, PVAAS models use all available testing data for each student. This minimizes the impact of a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

PVAAS in Practice

Each value-added estimate has an associated standard error, which is a measure of uncertainty that depends on the quantity and quality of student data associated with that value-added estimate. The standard error can help indicate whether a value-added estimate is significantly different from the growth standard. For PVAAS 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. These definitions then map to growth color indicators in the reports themselves.

Figure 6 below shows visual examples of each category. The green line represents the expected growth. The solid black line represents the range of values included in the growth measure plus and minus *one* standard error. The dotted black line extends the range of values to the growth measure plus and minus *two* standard errors. If the dotted black line is completely above expected growth, then there is significant evidence that students made more than expected growth, which represents the 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 6: Visualization of Growth Color Categories with Expected Growth, Growth Measures, and Standard Errors



Topics Related to the Value-added Modeling Approach Itself

Topic: PVAAS is based on a "black box" methodology.

The PVAAS methodologies and algorithms are published and have been in the open literature for almost 20 years. For those interested in learning more about the statistical models used in EVAAS reporting, including PVAAS in Pennsylvania, the following references are useful:

- On the **SAS EVAAS Statistical Models specific to PVAAS**: SAS Institute Inc. *SAS® EVAAS Statistical Models and Business Rules of PVAAS Analyses.* (Available online at https://pvaas.sas.com/support/PVAAS-Technical-Documentation.pdf).
- On the **Tennessee Value-Added Assessment System**: Millman, Jason, ed. *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* Thousand Oaks, CA: Corwin Press, 1997.

PVAAS in Theory

Although the modeling approach underlying PVAAS can be widely understood at a high-level, many details of the implementation require a strong background in statistics and value-added modeling to understand fully. This statistical rigor is necessary to provide reliable growth estimates. More specifically, the PVAAS models attain their reliability by using advanced techniques to address critical issues related to working with student testing data, such as students with missing test scores and the inherent measurement error associated with any test score.

These details have been published and made available to a community of experts for review and critique. Through this process, the PVAAS modeling has been sufficiently understood such that value-added experts and researchers have replicated the models for their own analyses. In doing so, they have validated and reaffirmed the appropriateness of the PVAAS modeling, and many of the early concerns were later assuaged through subsequent research and understanding. The references below include recent studies by statisticians from the RAND Corporation, a non-profit research organization:

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

PVAAS in Practice

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

- The growth standard methodology (also known as the multivariate response model or MRM) used in value-added analyses is a multivariate, longitudinal, linear mixed model. In other words, it is conceptually a multivariate repeated-measures ANOVA model. The growth standard methodology is used when scores are scaled or transformed so that the difference between two scores is meaningful and when there is clear "before" and "after" assessments in which to form a reliable gain estimate. In Pennsylvania, this model is used for PSSA Mathematics and ELA, grades 4–8.

- The predictive methodology (also known as the univariate response model or URM) used in value-added analyses is conceptually an analysis of covariance (ANCOVA) model. The predictive methodology is used when the test data do not meet the requirements for MRM analyses as stated above. In Pennsylvania, this model is used in subjects that are not tested in consecutive grades, such as PSSA Science and Keystones. For more details, please visit the PVAAS e-learning modules which can be found at: https://pvaas.sas.com/learningModules.html.

Topic: The PVAAS methodology is too complex; a simpler approach to measuring growth would provide better information to educators.

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

Value-added estimates based on simple calculations are often correlated with the type of students served by the educators. Such models often unfairly disadvantage educators serving students with a history of lower achievement and unfairly advantage educators serving students with a history of higher achievement.

Furthermore, it is not necessary to have an in-depth understanding of the modeling underlying PVAAS to use it to make effective decisions. Just as most people can use weather forecasts to make everyday decisions without in-depth knowledge on how meteorologists determine them, educators can use data provided by PVAAS reporting to guide their practice. With the PVAAS web application, educators have a wealth of reports that go beyond a single estimate of student growth and assist in identifying accelerants and impediments to student learning. These reports have been designed to be approachable and useful for educators with or without advanced statistical backgrounds.

PVAAS in Theory

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

- **How to accommodate team teaching** or other scenarios where more than one instructor has responsibility for a student's learning.
- **How to dampen the effects of measurement error** inherent in all student assessments because the tests themselves are estimates of student knowledge and not an exact measurement.
- **How to accommodate students with missing test scores** without introducing major biases by eliminating the data for students with missing scores, using overly simplistic imputation procedures, or using very few test scores for each student.
- **How to use all the longitudinal data for each student when the historical data are not on the same scale.**
- **How to use historical data when testing regimes have changed over time** to provide educational policymakers flexibility.

PVAAS modeling approaches address these concerns to provide reliable estimates of student growth. In particular:

- **PVAAS value-added measures are based on multiple years and assessed content areas of performance data (rather than one prior test score) to determine students' academic growth in LEAs/districts, schools, and classrooms.** The inclusion of multiple years of data from multiple subjects for each individual student adds to the protection of an educational entity from misclassification in the value-added analysis. More specifically, using that much data at the individual student level can dampen the effect of measurement error, which is inherent in any test score and in all value-added or growth models.
- **PVAAS value-added measures are sophisticated and robust enough to include students with missing data.** Since students with a history of lower achievement are more likely to miss tests than students with a history of higher achievement, the exclusion of students with missing test scores can introduce selection bias, which would disproportionately affect educators serving those students.
- **PVAAS value-added measures provide estimates whether, on average, the students fell below, met, or exceeded the established expectation for improvement in a particular grade/subject.** Assessing the impact at the group level, rather than on individual students, is a more statistically reliable approach due to the issues with measurement error.
- **PVAAS value-added measures account for the amount of evidence (standard error) when determining whether an educational entity is decidedly above or below the growth standard as defined by the model.** Any model based on assessment data relies on estimates of student learning, and it is important that any value-added measure account for the amount of evidence in the growth measure when providing estimates.
- **PVAAS value-added models are sophisticated enough to accommodate different tests or changes in testing regimes.** This provides educators with additional flexibility. First, they can use more tests, even if they are on different scales. Second, they can continue to provide reporting when tests change.

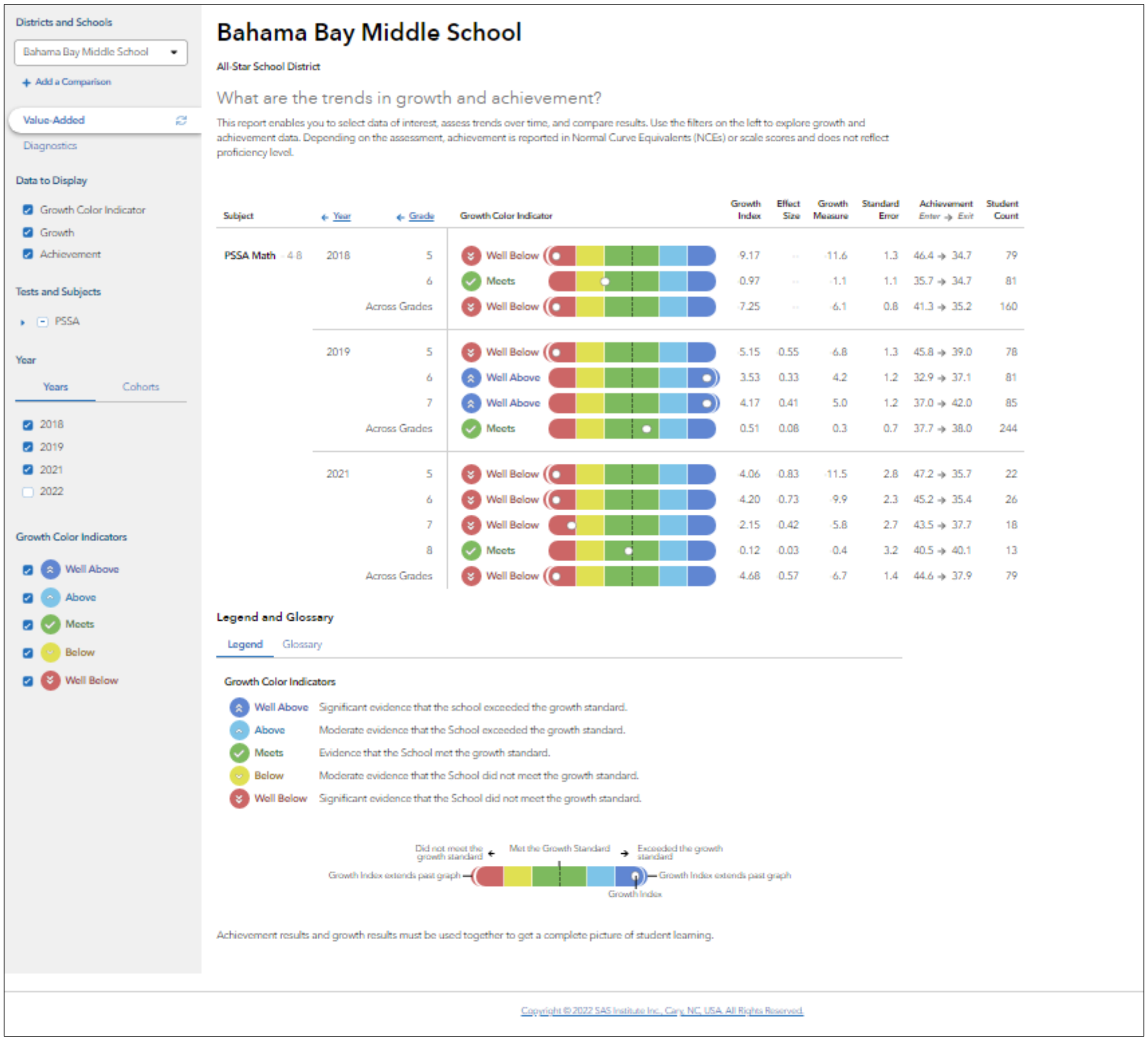
PVAAS in Practice

Although the statistical approach is robust and complex to safeguard against the issues previously described, the reports in the PVAAS web application are easy to understand. Figure 7 below shows an example value-added report as displayed in the web application. Provided by subject, grade, and year, the value-added estimates are color-coded for quick understanding:

- Dark blue or light blue indicates that students with an LEA/district, school, or teacher exceeded the growth standard.
- Green indicates that students with an LEA/district, school, or teacher met the growth standard.
- Yellow or red indicates that students with an LEA/district, school, or teacher did not meet the growth standard.

Educators and administrators can identify their strengths and opportunities for improvement at a glance. The reporting is interactive, so that authorized users can drill down to access diagnostic reports for students by achievement level, individual student-level projections to achievement, and other reports. Educators have a comprehensive view of past practices as well as tools for current and future students. Thus, educators benefit from the rigor of the PVAAS models by gaining insight in an accessible and non-technical format. PVAAS Value-Added reports are customized for Pennsylvania reporting and preferences. For more details about the reporting, please visit the PVAAS e-learning modules which can be found at: https://pvaas.sas.com/learningModules.html.

Figure 7: Sample PVAAS School Value-Added Report



Topic: How can PVAAS accommodate the realities of today's classroom?

The instruction that students receive from educators can be much more complex than one teacher for a given subject and grade. In today's classroom, there might be team teaching, pull out or push in programs, lab sessions, English Language Development (ELD) instruction, or countless other ways that more than one instructor is responsible for a student's learning in a particular subject and grade. It is important to capture such contributions in teacher value-added reporting, and PVAAS does just that.

PVAAS in Theory

The statistical modeling underlying PVAAS uses a robust approach that can account for team teaching or other scenarios where more than one instructor is responsible for a student's learning in a particular subject and grade. If just one teacher is responsible for a student's learning, that student is weighted fully in that teacher's Value-Added report. If more than one teacher is responsible for a student, then the student is weighted in each Teacher Value-Added report according to the percentage of instructional responsibility that the teacher has. A teacher's Value-Added report reflects all the students linked to them, and it considers the appropriate weighting.

PVAAS in Practice

The weighting itself is captured by the roster verification process available through the PVAAS web application. This application allows teachers and their administrators to review and modify the list of students linked to them. This step provides an important measure of verification and validation for accurate student-teacher linkages. A sample screenshot using demonstration data is in Figure 8 below.

Any protocols and policies on which educators to include in roster verification and how to assign the percentage of instructional responsibility to educators were determined by the PDE and based on results from the pilot study.

Figure 8: Sample Roster Verification for a Teacher

Add Student | Remove Student

Status: Teacher Is Verifying

	Student	Student ID	Instructional Responsibility			Edits
			Student + Teacher Assignment	Full or Partial % of Instruction	Total	
			Select All	Select All		
1	ACEVEDO, OSIEL	428346753	50 %	50 %	25.0%	
2	AHSAN, CHANDANI	576916773	100 %	100 %	100.0%	
3	BAILEY, JASPER	419842348	100 %	100 %	100.0%	
4	BEAN, CASE	191188873	100 %	100 %	100.0%	
5	BECK, MARISSA	853132004	100 %	100 %	100.0%	
6	BLACK, RYAN	445205625	100 %	100 %	100.0%	
7	BLEVINS, CAMDEN	833364450	100 %	80 %	80.0%	
8	BOND, SHANDA	737578609	100 %	50 %	50.0%	
9	BOWLING, KATHY	566440733	100 %	100 %	100.0%	
10	BOWMAN, GILBERT	282738426	70 %	100 %	70.0%	
11	BROWN, MYKAYLA	271385113	20 %	100 %	20.0%	
12	BRUCE, EDEN	718933555	100 %	100 %	100.0%	
13	BUSTOS, MARA	232605258	90 %	100 %	90.0%	

Topic: Are teachers of small classes disadvantaged with PVAAS?

The PVAAS Teacher Value-Added report provides a value-added (or growth) measure as well as a standard error. The standard error provides information about the amount of evidence or data used in the growth measure. The two metrics are used together to assess whether there is significant or moderate evidence that the teacher’s group of students met, exceeded, or fell short of the growth standard. The standard error is based on the number of students linked to the teacher as well as the variability in those students’ test scores. Although there might be concern that teachers of small classes are disadvantaged by PVAAS, they are protected by using a value-added estimate and standard error together.

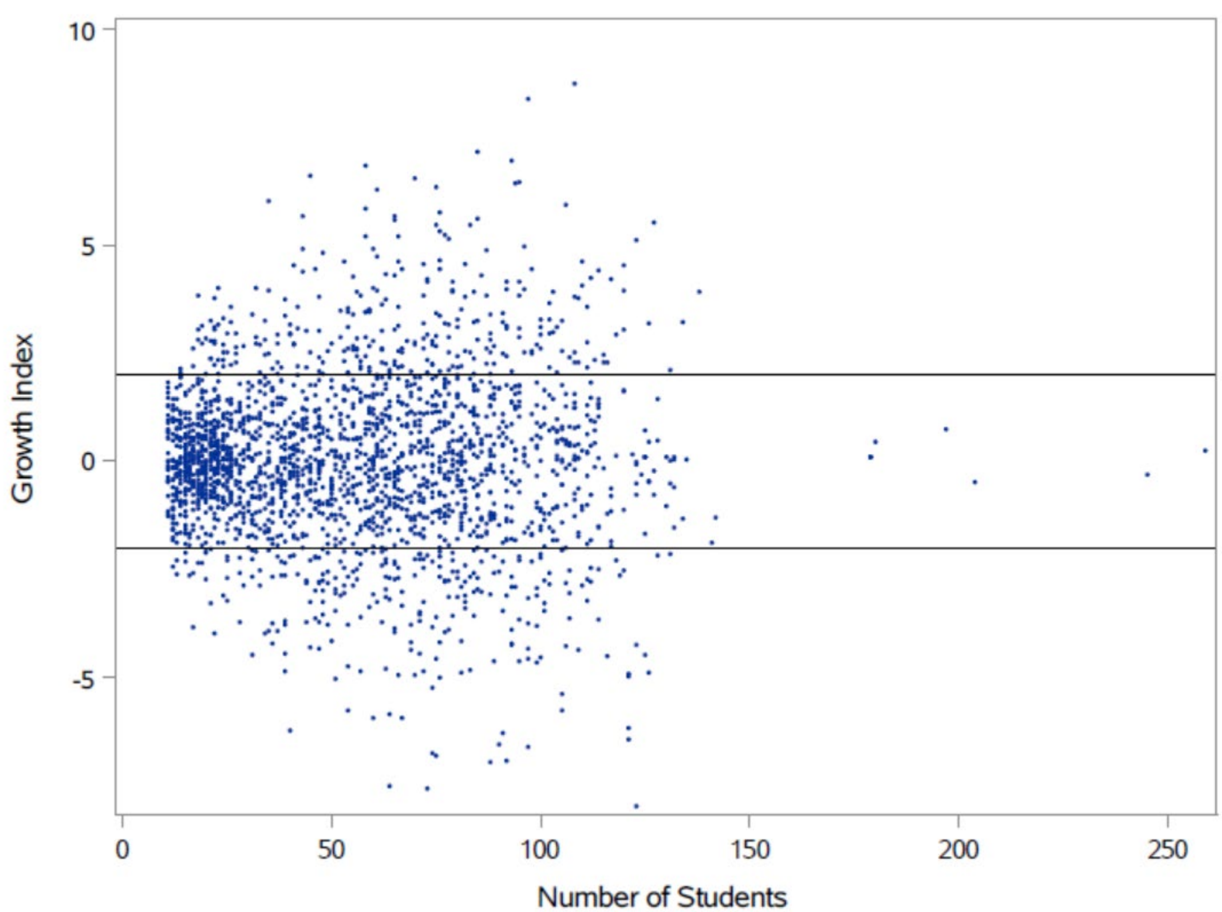
PVAAS in Theory

Students in all class sizes have the ability to show growth, and the standard error simply provides a confidence band around each estimate. With a smaller amount of data (meaning fewer students), there is less evidence in each estimate, so the standard error tends to be larger than teachers linked to a large number of students. However, although teachers of small classes *might* have larger standard errors than other teachers, they are also more likely to have a larger gain—either positive or negative. Thus, the two metrics even out, and teachers of small classes are not disadvantaged.

PVAAS in Practice

Actual data might be the most readily apparent way to demonstrate that small classrooms show similar growth as large classrooms. The graph in Figure 9 plots the number of students used in each teacher's PVAAS Value-Added report against the teacher growth index (the value-added estimate divided by its standard error) for PSSA Mathematics in grade 7 in 2023. Each dot represents one teacher, and verified rosters were used where available. The graph demonstrates that teachers serving both small and large numbers of students can show both high and low growth, as measured by PVAAS. Although current state policy requires that teachers are linked to at least eleven individual student scores to receive a teacher report, the graph below shows that even teachers of very small classrooms are not disadvantaged with a sophisticated value-added approach, like PVAAS. In the graph below, the actual correlation between the growth index and number of students is -0.001, which is negligible.

Figure 9: Pennsylvania Growth Index Versus Number of Students Linked to Teacher



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

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

PVAAS in Theory

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

To explore the reliability of teachers’ value-added estimates, SAS reviewed results from the past two decades using data from another state that uses methodology similar to PVAAS and found that:

- **Teachers with high value-added are likely to continue yielding high value-added.** Teachers identified as having students who exceed expected growth after their first three years of teaching were extremely likely to have similar growth with their students three years into the future (about 95% were either average or above average in their students’ growth).
- **Teachers with lower value-added might improve over time.** For the teachers identified as having students who do not meet expected growth based on three-year estimates, approximately half of them continue to have students with similar growth three years later.

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

PVAAS in Practice

For individual subjects and grades in 2023, less than 1% of teachers (.01%or six teachers) moved from the highest growth designation (based on significant evidence of exceeding the growth standard) to the lowest growth designation (based on significant evidence of not meeting the growth standard) from 2021-22 to 2022-23 based on the teachers’ composite value-added measures. In all individual subjects and grades, less than 3% move from the highest growth designation to the lowest growth designation.

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

The 2013 Measures of Effective Teaching (MET) study raised the same question of whether value-added data, in conjunction with other metrics like observational studies, could be used for high-stakes decisions and it concluded:

> The answer is yes, not because the measures are perfect (they are not), but because the combined measure is better on virtually every dimension than the measures in use now. There is no way to avoid the stakes attached to every hiring, retention, and pay decision. And deciding

> not to make a change is, after all, a decision. No measure is perfect, but better information should support better decisions.[5]

The MET study went on to encourage data practices to improve each measure, such as roster verification for student-teacher linkages and multi-year averages of teacher estimates, both of which are used for PVAAS teacher value-added reporting.

[5] Measures of Effective Teaching Project. "Ensuring Fair and Reliable Measures of Effective Teaching," (Seattle, WA: Bill and Melinda Gates Foundation, 2013), 13.