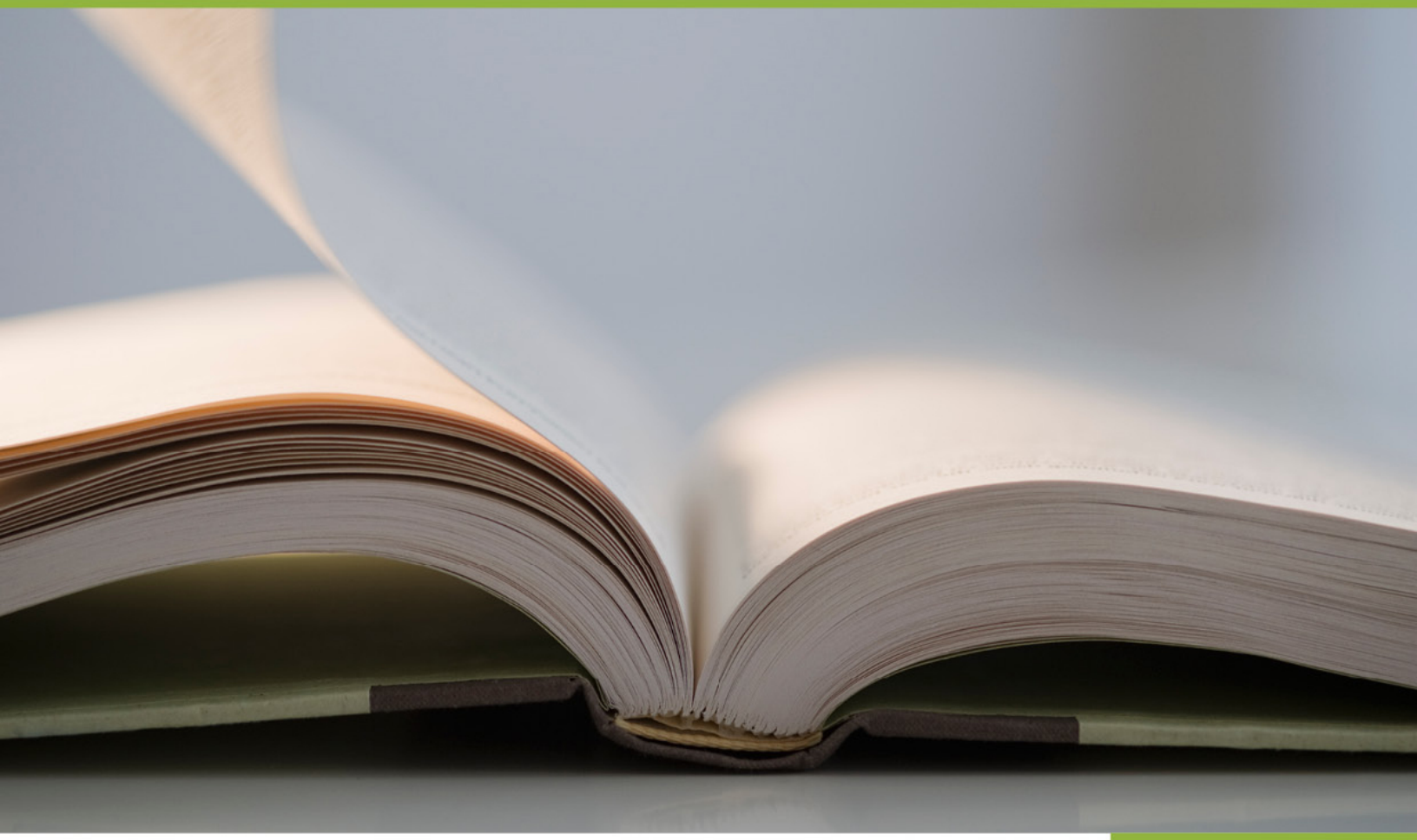
SAS® EVAAS®

Misconceptions about PVAAS Reporting for Pennsylvania



THE POWER TO KNOW.

Introduction **1**

Misconceptions related to the student population served by educators **1**

It is harder to make growth with students from certain demographic or socioeconomic backgrounds. 1

If students are already high achieving, it is harder to show growth. 5

PVAAS should always indicate growth if the percentage of students scoring Proficient or above increased since last year. 7

PVAAS cannot measure the progress of students with missing data or highly mobile students. 8

Misconceptions related to the test used in value-added modeling **9**

PVAAS reporting is not reliable or valid since it is based only on the PA state assessments. 9

Misconceptions related to the value-added modeling approach itself **10**

PVAAS is based on a "black box" methodology. 10

The PVAAS methodology is too complex; a more simple approach to measuring district, school and teacher effectiveness would provide better information to educators. 11

How can PVAAS accommodate the realities of today's classroom? 14

Are teachers of small classes disadvantaged with PVAAS? 15

Teacher value-added estimates are not reliable enough to be used in high-stakes decisions. 17

EVPAPDE-1005-03OCT2016. Copyright © 2016 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.

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 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 school district and school based on the statewide Pennsylvania System of School Assessments (PSSA) and Keystones by means of a secure Web application.

The value-added estimates provided by PVAAS are based on a robust and reliable methodology. This important approach overcomes many critical statistical issues related to using standardized tests to assess student progress, and mitigates concerns about fairness. This document addresses the following misconceptions about PVAAS reporting.

Misconceptions related to the student population served by educators

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 is not typically the case for the most sophisticated value-added models, such as those used for PVAAS in the Commonwealth of Pennsylvania. This is because, regardless of the specific modeling approach that is used in Pennsylvania, the PVAAS approach uses all available testing history for each individual student and does not exclude students who have missing test data. In essence, each student serves as his or her own control, and to the extent that SES/DEM influences persist over time, these influences are already represented in the student's data.

PVAAS in Theory

As a 2004 Ed Trust study stated, specifically with regards to the SAS EVAAS modeling, which is the approach used in Pennsylvania's 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]

A 2007 paper by RAND researchers J.R. Lockwood and Dan McCaffrey explicitly verified the models used for PVAAS district, school and teacher reporting when they wrote:

> *William Sanders, the developer of the TVAAS model, has claimed that jointly modeling 25 scores for individual students, along with other features of the approach is extremely effective at*

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

> *purging student heterogeneity bias from estimated teacher effects… The analytic and simulation results presented here largely support that claim.*[2]

An economist-based perspective by UCLA researchers Pete Goldschmidt, Kilchan Choi, and Kyo Yamashiro provided a similar finding in their study comparing value-added models:

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

Ultimately, there may 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

While 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 may be the most readily apparent evidence. The figures below provide teacher-level data, and the results are similar to those for 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 five in 2014. Each dot represents one teacher, and verified rosters were used where available. Regardless of the student characteristics served by teacher, there is essentially 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 -0.139, which is weak.

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

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

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

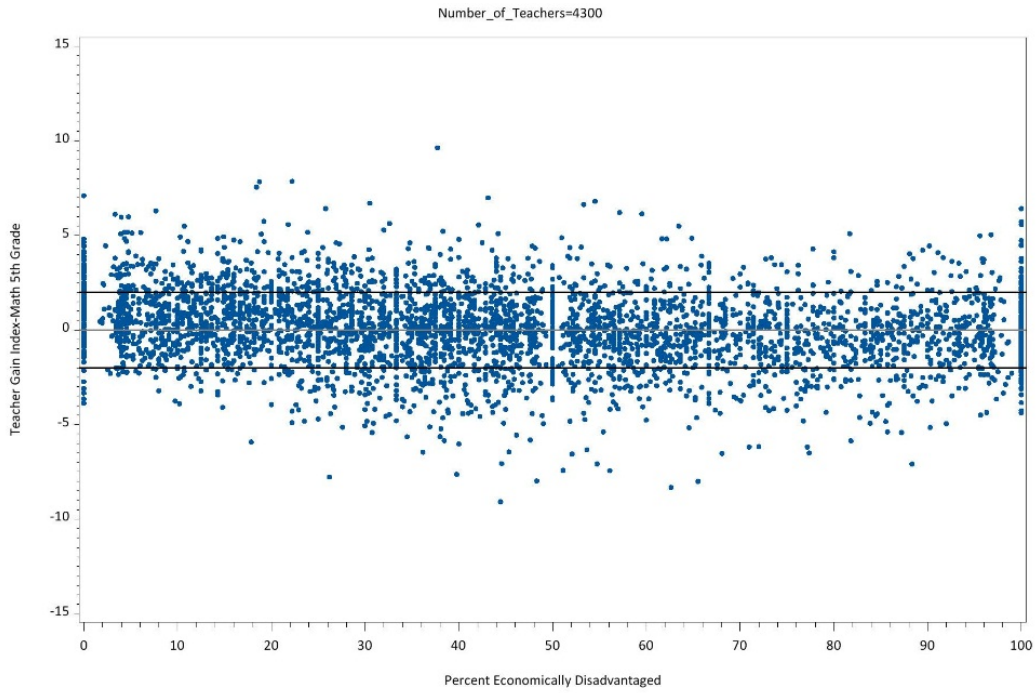


Figure 2 provides similar information for the percentage of minority students. Again, there is essentially no correlation to the growth index. In the graph below, the actual correlation between the growth index and percentage of students considered minority is -0.070, which is negligible.

Figure 2: Pennsylvania Growth Index V. Percent Tested Minority by Teacher

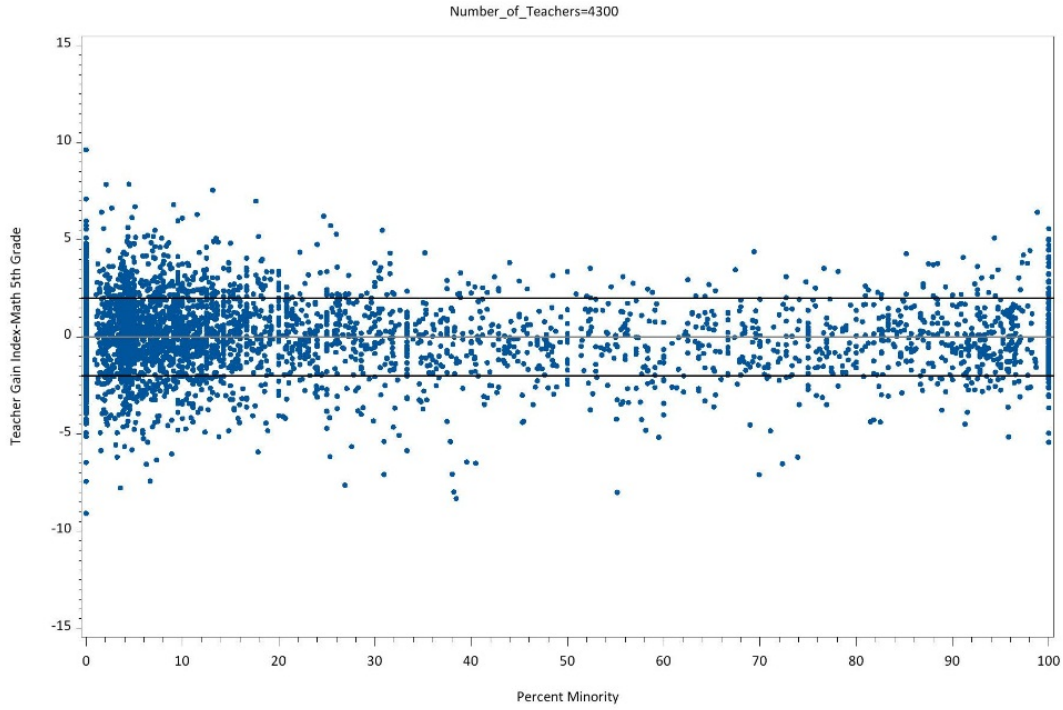


Figure 3 provides similar information for the percentage of students considered Limited English Proficient (LEP), and there is essentially no correlation to the growth index. In the graph below, the actual correlation between the growth index and percentage of students testing as LEP is -0.006, which is negligible.

Figure 3: Pennsylvania Growth Index V. Percent Tested LEP by Teacher

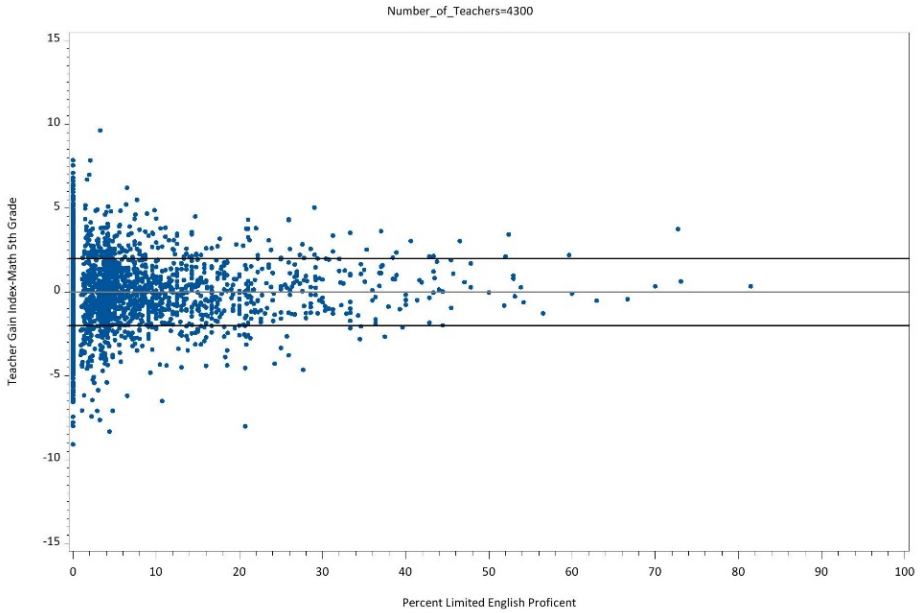
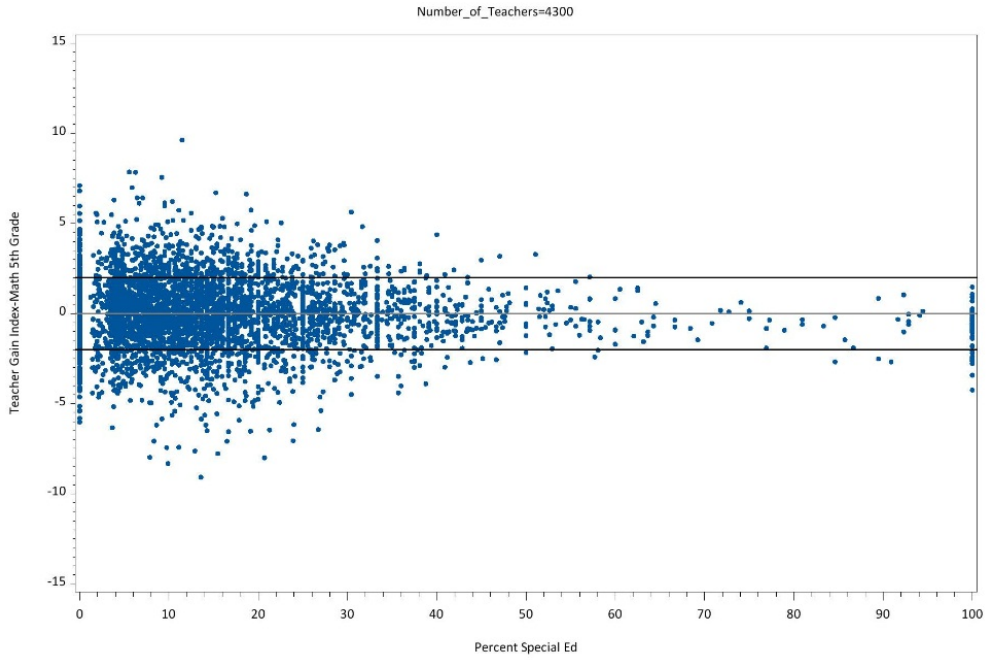


Figure 4 provides similar information for the percentage of students considered special education, and there is essentially no correlation to the growth index. In the graph below, the actual correlation between the growth index and percentage of students testing as special education is -0.096, which is negligible.

Figure 4: Pennsylvania Growth Index V. Percent Tested Special Education by Teacher



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

Educators serving high-achieving students 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 progress 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 students made one year's worth of growth, which is about meeting expected performance based on a specific group's prior academic performance.

Furthermore, while 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 measure the growth of high-achieving students.

In fact, any test that is used in PVAAS analyses must meet three criteria, and state assessments meet these criteria:

- Aligned to curriculum standards.
- Reliable and valid.
- Demonstrate sufficient stretch at the extremes.

PVAAS is fair not only to districts, schools and teachers serving high-achieving students; it is fair to the students themselves. The modeling that underlies PVAAS takes into account the progress of all students, regardless of their entering achievement, and the reporting shows whether the curriculum and instruction target appropriately to these students. High-achieving students may require enrichment work in the same way that low-achieving students may need remediation to make sufficient progress.

PVAAS in Practice

Actual data may be the most readily apparent evidence. The graph in Figure 5 plots the average entering 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 five in 2014. Each dot represents one teacher, and verified rosters were used where available. Regardless of the teacher's student achievement, there is essentially 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 0.131, which is weak. District and school value-added plots are similar to the teacher one shown below. This shows that high achieving students can—and do—show growth through PVAAS, and that educators are neither advantaged nor disadvantaged by the achievement level of their students.

Figure 5: Pennsylvania Growth Index V. Average Achievement by Teacher

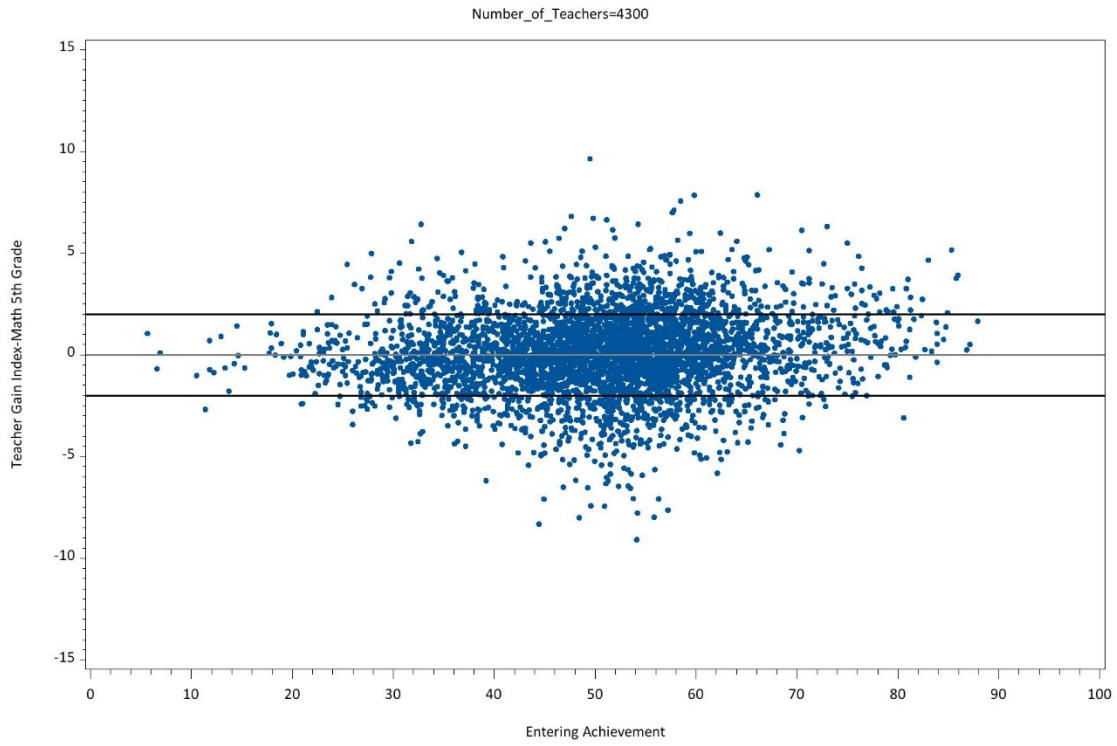
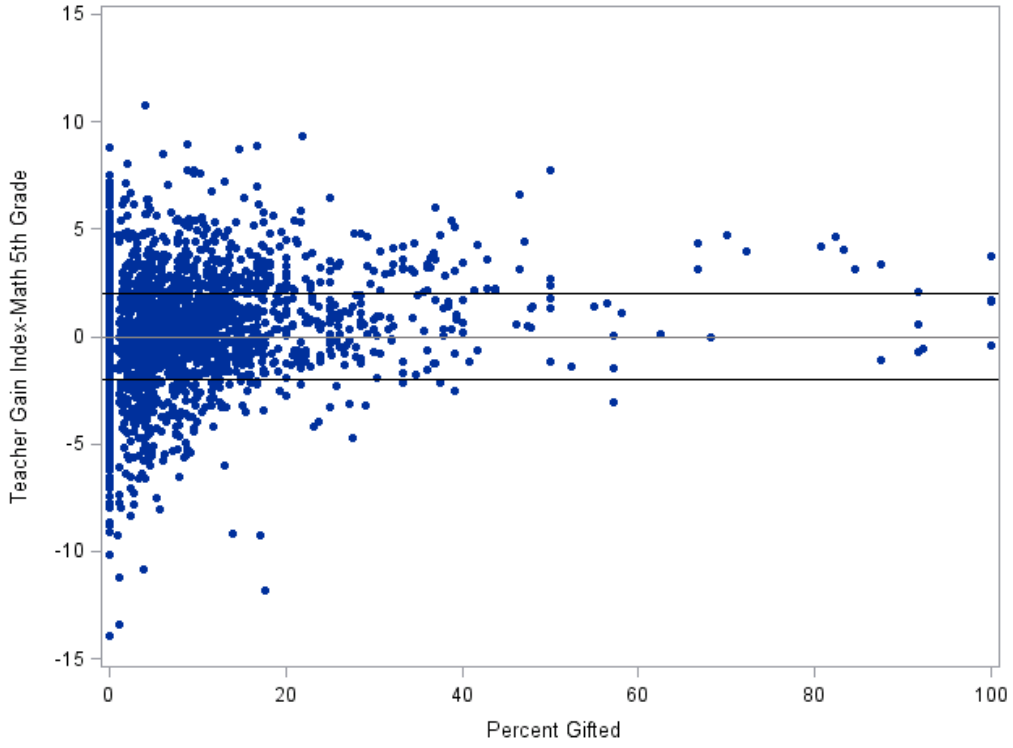


Figure 6: Pennsylvania Growth Index V. Percent Gifted by Teacher



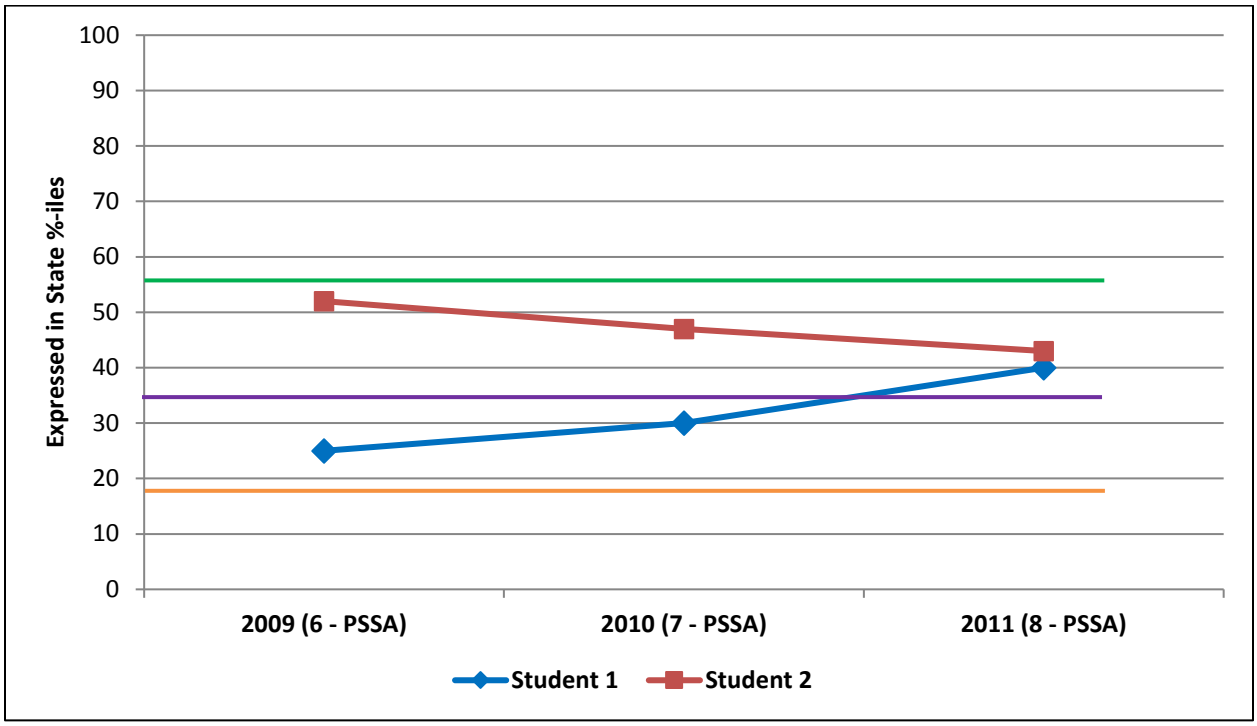
PVAAS should always indicate growth if the percentage of students scoring Proficient or above increased since last year.

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

Imagine the scenario below. The mathematics achievement level of Student 1 is represented by the line with the blue diamonds, and that of Student 2 is represented by the line with the red squares. The achievement level of Student 1 has steadily increased over time while the achievement level of Student 2 has steadily decreased over time. From seventh to eighth grade, Student 1 moved from the Basic to Proficient performance category. From seventh to eighth grade, Student 2 maintained his position in the Proficient performance category, although his achievement level has gone down.

Figure 7: Student Testing History in PSSA Mathematics for Student 1 and Student 2



Just by considering the number of students who have scored Proficient, assuming all other students have maintained the same performance categories, the number of students has increased with the addition of Student 1. However, this does not take into account that Student 2's achievement level is steadily decreasing over time. A more subtle approach is required that considers the growth of all students, regardless of their achievement level.

PVAAS in Practice

PVAAS does not measure students' progress based on the number or percentage of students who tested Below Basic, Basic, Proficient, or Advanced, as compared to previous years. PVAAS detects these subtle

changes in progress, even within performance levels. As a result, educators are recognized when they make progress with students at/above proficiency *and* below/not yet proficient. This can be very encouraging to districts, schools and teachers serving low-achieving students, who might not otherwise be recognized for their students' progress.

PVAAS cannot measure the progress of students with missing data or highly mobile students.

PVAAS value-added analyses provide reliable and valid estimates of the effectiveness of 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 progress is representative of the students actually served by districts, schools, and teachers.

PVAAS in Theory

Highly mobile students are more likely to be low-achieving students, and it is important to include these students to avoid selection bias, which could provide misleading growth estimates to districts, schools, and teachers. While more simplistic value-added or growth estimates may require that students have the same set of prior test scores or that students have all prior test scores, this often has the result of excluding mobile student populations, and this would disproportionately affect educators serving those types of students.

Unlike simplistic approaches, 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 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 the Pennsylvania Department of Education's request, students are excluded for other reasons, such as first-year ELL status or not meeting the full academic year (FAY) requirement.

[4] Wright, S. P. (2004). "Advantages of a Multivariate Longitudinal Approach to Educational Value-Added Assessment Without Imputation." Paper presented at National Evaluation Institute, online at https://pvaas.sas.com/support/EVAAS-AdvantagesOfAMultivariateLongitudinalApproach.pdf.

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

Misconceptions related to the test used in value-added modeling

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

Educators may be concerned that value-added reporting relies on the use of standardized tests, which have limitations themselves. Perhaps they feel that the test does not correlate well with the curriculum or that there isn't sufficient stretch to measure progress of very low- or high-achieving students. However, 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. SAS 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:

- Must be sufficiently aligned to curriculum standards.
- Must be reliable and valid.
- Must demonstrate sufficient stretch at the extremes.

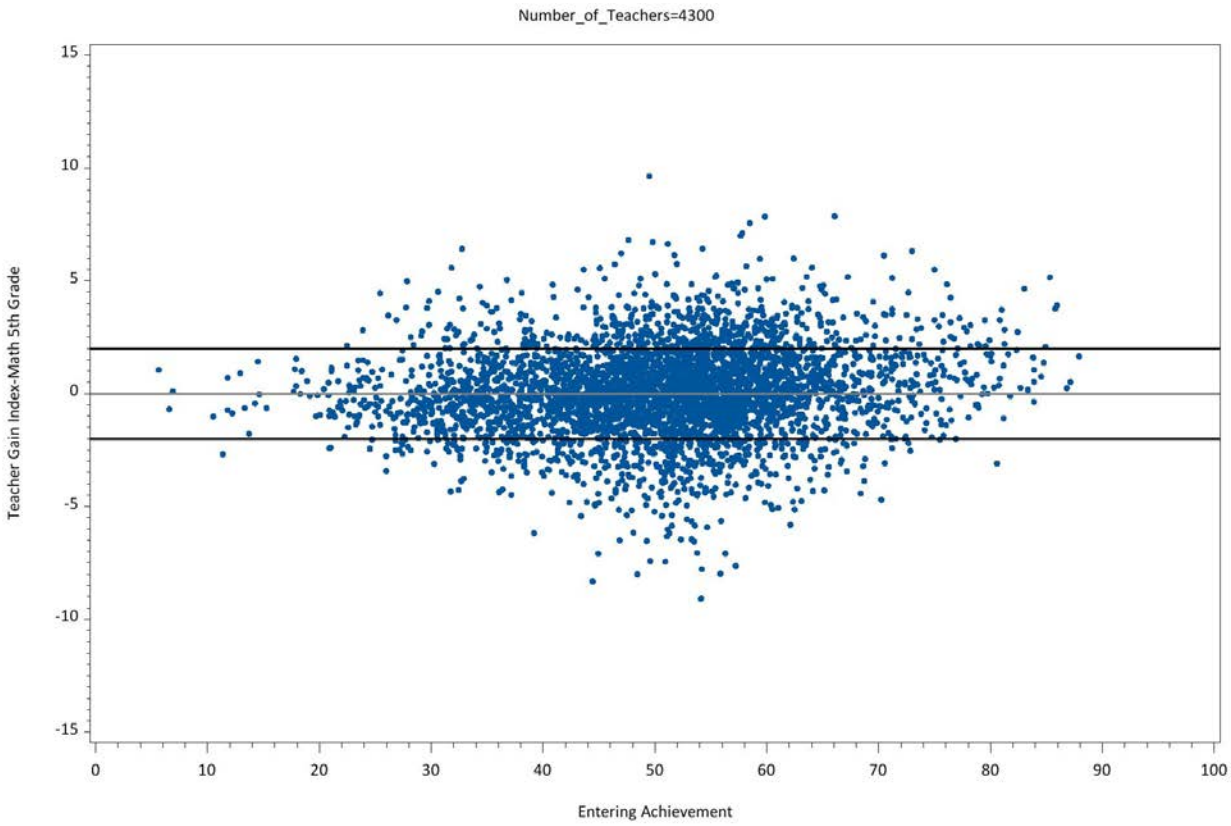
To date, Pennsylvania's state assessment have met these criteria. More specifically, EVAAS analyses verify that there are enough different scaled scores at the top and bottom of the scales to differentiate student achievement. EVAAS processing also analyzes the percentage of students scoring at the top 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 districts, schools, and teachers serving both high and low achieving students can show both high and low growth. This process is repeated every year.

Another common concern of educators is that they may 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 just represents 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.

PVAAS in Practice

Actual data may be the most readily apparent evidence. The graph in Figure 8 plots the average achievement for each teacher's students in Pennsylvania against the teacher growth index (the value-added growth measure divided by its standard error) for PSSA Mathematics in grade five in 2014. Each dot represents one teacher, and verified rosters were used where available. The graph demonstrates that teachers serving both high- and low-achieving students can show both high and low growth, as measured by PVAAS. District and school value-added plots are similar to the teacher plot shown below.

Figure 8: Pennsylvania Growth Index V. Average Achievement by Teacher



Misconceptions related to the value-added modeling approach itself

PVAAS is based on a "black box" methodology.

The PVAAS methodologies and algorithms are published and have been in the open literature for over 15 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 upon which Pennsylvania's reporting is based**: Wright, S.P., White, J.T., Sanders, W.L. and Rivers, J. C. (2010). "*SAS® EVAAS® Statistical Models*." March 25, 2010 (available at http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/sas-evaas-k12-statistical-models-107411.pdf.
- On the **Tennessee Value-Added Assessment System**: Millman, J. (ed.) (1997). "*Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?*"

PVAAS in Theory

While PVAAS reporting benefits from a robust modeling approach, this statistical rigor is necessary to provide reliable estimates. More specifically, the PVAAS models attain their reliability by addressing critical issues related to working with student testing data, such as students with missing test scores and the inherent measurement error associated with any test score.

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

  On the **insufficiency of simple value-added models**: McCaffrey, D. F., Han, B. and Lockwood, J. R. (2008). "From Data to Bonuses: A Case Study of the Issues Related to Awarding Teachers Pay on the Basis of the Students' Progress." Presented at *Performance Incentives*: *Their Growing Impact on American K-12 Education*, Feb. 28-29, 2008, National Center on Performance Incentives at Vanderbilt University.

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 multivariate response model (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 MRM is used when scores are scaled or transformed so that the difference between two scores is meaningful and when there are 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 four through eight.
- The univariate response model (URM) used in value-added analyses is conceptually an analysis of covariance (ANCOVA) model. The URM 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.

The PVAAS methodology is too complex; a more simple approach to measuring district, school and teacher effectiveness would provide better information to educators.

Although conceptually easy, the statistical rigor necessary to provide precise and reliable growth measures requires that several important analytical problems be addressed when analyzing longitudinal student data, which is critically important in any reporting used for educator evaluations.

In short, a simple gain calculation does not provide a reliable estimate of educators' effectiveness. Value-added estimates based on simple calculations are often correlated with the type of students served by the educators, rather than the educator's effectiveness with those students. Such models often unfairly disadvantage educators serving low-achieving students and unfairly advantage educators serving high-achieving students.

However, it is not necessary to be a statistician to understand the educational implications of EVAAS reporting. With the PVAAS Web application, educators have a wealth of reports that go beyond a single estimate of effectiveness and assist in identifying accelerants and impediments to student learning.

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**, which is inherent in all student assessments because the tests themselves are estimates of student knowledge, not an exact measurement.
- **How to accommodate students with missing test scores** without introducing major biases by eliminating the data for students with missing scores, using overly simplistic imputation procedures, or using very few test scores for each student.
- **How to exploit all of the longitudinal data for each student when all of the historical data are not on the same scale.**
- **How to use historical data when testing regimes have changed over time** to provide educational policymakers flexibility.

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

- **PVAAS value-added measures are based on all of a student's previous years' performance data on an assessment instrument (rather than one prior test score) to determine students' academic progress in 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 all available data at the individual student level can dampen the effect of measurement error, which is inherent in any test score and in all value-added or growth models.
- **PVAAS value-added measures are sophisticated and robust enough to include students with missing data.** Since low-achieving students are more likely to miss tests than high-achieving students, the exclusion of students with missing test scores can introduce selection bias, which would disproportionately affect educators serving those students.
- **PVAAS value-added measures provide estimates whether, on average, the students fell below, met, or exceed the established expectation for improvement in a particular grade/subject.** Assessing the impact at the group level, rather than on individual students, is a more statistically reliable approach, due to the issues with measurement error.
- **PVAAS value-added measures take into account the measures of uncertainty (standard error) when determining whether an educational entity is decidedly above or below expected progress, as defined by the model.** Any model based on assessment data relies on estimates of student learning, and it is important that any value-added measure take into account the inherent uncertainty when providing estimates.
- **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 differing scales. Second, they can continue to provide reporting when the tests change, as is the case in Pennsylvania.

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

- On the **choice of a complex value-added model**: McCaffrey, D. F., Han, B. and Lockwood, J. R. (2008). "Value-Added Models: Analytic Issues." Prepared for the National Research Council and the National Academy of Education, Board on Testing and Accountability Workshop on Value-Added Modeling, Nov. 13-14, 2008, Washington D.C.
- On the **advantages of the longitudinal, mixed model approach**: Lockwood, J.R. and McCaffrey, D.F. (2007). "Controlling for Individual Heterogeneity in Longitudinal Models, with Applications to Student Achievement." Electronic Journal of Statistics, Vol. 1, 223-252.
- On the **insufficiency of simple value-added models**: McCaffrey, D. F., Han, B. and Lockwood, J. R. (2008). "From Data to Bonuses: A Case Study of the Issues Related to Awarding Teachers Pay on the Basis of the Students' Progress." Presented at Performance Incentives: Their Growing Impact on American K-12 Education, Feb. 28-29, 2008, National Center on Performance Incentives at Vanderbilt University.

It is not just the models that use a sophisticated and robust process; from the moment that PSSA and Keystones data arrive at SAS, they are subjected to a rigorous review to verify that these data are appropriate for value-added analyses. SAS uses a sophisticated process in tracking students over time, which accommodates many common data problems at the individual student level, such as missing test scores, duplicate scores, or changing student data.

PVAAS in Practice

Although the statistical approach is robust and complex, the reports in the PVAAS Web application are easy to understand. Provided by subject, grade, and year, the value-added estimates are color-coded for quick understanding: blue indicates that students with a district, school or teacher made more than the expected progress; green indicates that students with a district, school or teacher made about the expected progress; and yellow or red indicates that students with a district, school or teacher made less than the expected progress. Educators and administrators can identify their strengths and opportunities for improvement at a glance. The reporting is interactive, so that authorized users can drill down to access diagnostic reports for students by subgroup or achievement level, individual student-level projections, and other reports. Educators have a comprehensive view of past practices as well as tools for current and future students. Thus, educators benefit from the rigor of the PVAAS models by gaining insight in an accessible and non-technical format. The PVAAS value-added reports are customized for Pennsylvania reporting and preferences, but the sample PVAAS school report below illustrates how PVAAS reports can be user-friendly and do not require sophisticated statistical knowledge.

Figure 9: Sample PVAAS School Value-Added Report

Report: School Value Added **Test:** PSSA
School: Hound Middle School **Subject:** Reading/ELA
District: East Suburb School District
Year: 2016

LEARN HOW TO USE THIS REPORT (Flash required)
Math/Reading 4-8 | Science, Writing, Keystone Exams

Estimated School Growth Measure				
Grade	6	7	8	Growth Measure over Grades Relative to Standard for PA Academic Growth
Standard for PA Academic Growth	0.0	0.0	0.0	
2014 Growth Measure	-4.5 R	-1.7 G	7.3 DB	0.4 G
Standard Error	1.4	1.7	1.1	1.0
2015 Growth Measure	-3.0 Y	-1.0 G	0.6 G	-1.2 Y
Standard Error	1.9	1.6	2.2	1.1
2016 Growth Measure	5.2 DB	2.5 LB	0.6 G	2.7 DB
Standard Error	1.2	1.7	2.1	1.0
3-Yr-Avg Growth Measure	-0.8 G	-0.0 G	2.8 DB	0.7 LB
Standard Error	0.9	1.0	1.0	0.6

Estimated School Avg Achievement			
Grade	6	7	8
State NCE Average	50.0	50.0	50.0
2013 Avg Achievement	51.5	47.3	41.9
2014 Avg Achievement	48.1	50.5	52.6
2015 Avg Achievement	50.1	47.6	52.0
2016 Avg Achievement	42.3	52.5	46.3

DB	Significant evidence that the school exceeded the standard for PA Academic Growth
LB	Moderate evidence that the school exceeded the standard for PA Academic Growth
G	Evidence that the school met the standard for PA Academic Growth
Y	Moderate evidence that the school did not meet the standard for PA Academic Growth
R	Significant evidence that the school did not meet the standard for PA Academic Growth

Achievement results (PA state assessments) and growth results (PVAAS) must be used together to get a complete picture of student learning.
PA Academic Growth for Math & Reading/ELA:
The standard for PA Academic Growth is met when the student group maintains their relative achievement level from one year to the next. (SY12-13/SY13-14-PSSA Reading; SY14-15-PSSA ELA)

To view additional reports, click on the underlined numbers or words.

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 may be team teaching, pull out or push in programs, lab sessions, English as a Second Language 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 take into account 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 him or her 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 9 below.

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

Figure 10: 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%	

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 is a measure of uncertainty, and the two metrics are used together to assess whether there is significant or moderate evidence that the teacher's students have made more or less than the expected growth. The standard error is based on the number of students linked to the teacher as well as the variability in those students' test scores. While there may be concern that teachers of small classes are disadvantaged by PVAAS, they are actually protected by using a value-added estimate and standard error together.

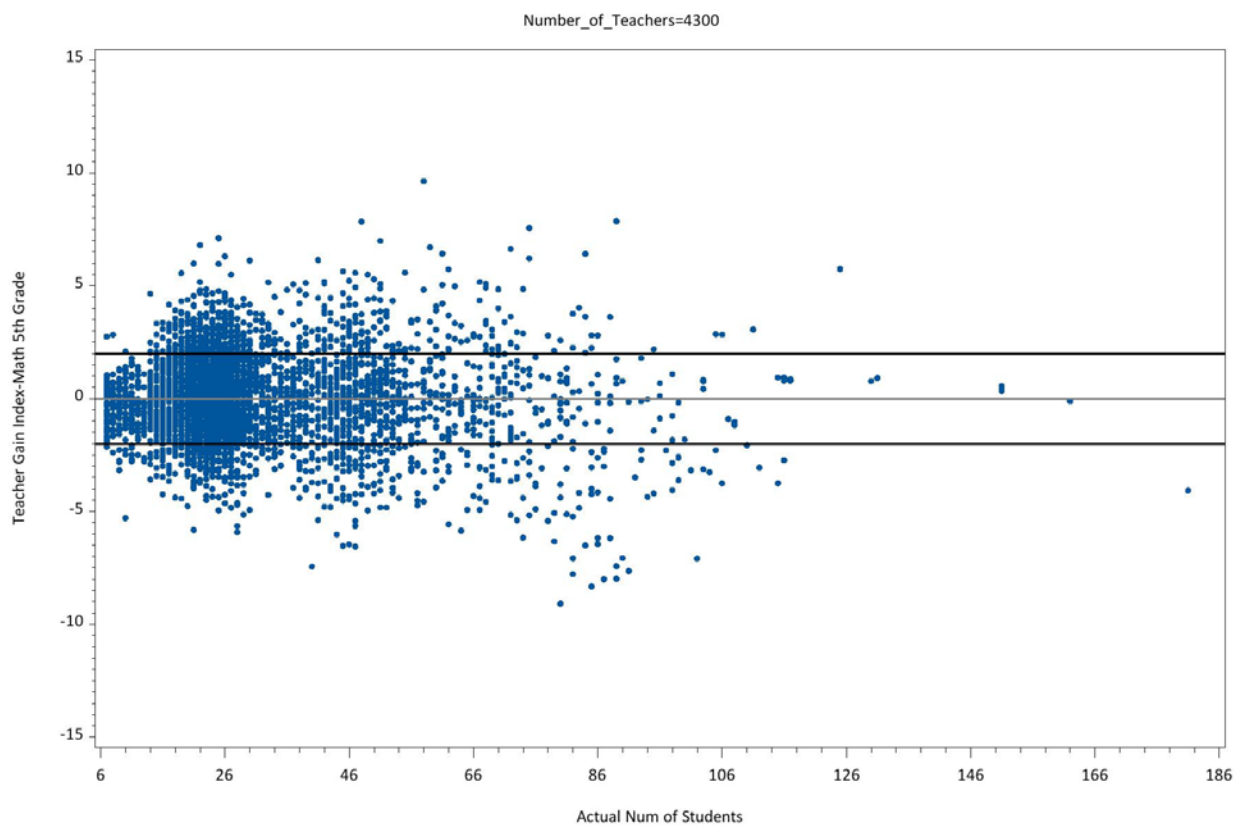
PVAAS in Theory

Students in all class sizes have the ability to show growth, and the standard error simply provides a confidence band around each estimate. With smaller amount of data (meaning, fewer students), there is less certainty around each estimate, so the standard error tends to be larger than teachers linked to a large number of students. However, while teachers of small classes *may* 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 may be the most readily apparent evidence. The graph in Figure 11 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 five in 2014. 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. While current state policy requires that teachers are linked to at least eleven students in order to receive a teacher report, the graph below shows that even teachers of very small class rooms 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 0.009, which is negligible.

Figure 11: Pennsylvania Growth Index V. Number of Students Linked to Teacher



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 very robust statistical approach and report a multiple-year average, whenever available. The approach provides very reliable teacher estimates, which educators can use for a variety of educational and policy decisions.

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 may be more effective with one group than another. Also, some teachers may improve, or worsen, in their effectiveness over time. However, the presence of strong reliability indicates that teachers' value-added estimates are related to their consistent skills and are not generated primarily from a random component.

SAS reviewed value-added estimates over 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 highly effective after their first three years of teaching were extremely likely to remain effective three years into the future (about 95% were either average or above average in effectiveness).
- **Teachers with lower value-added may improve over time.** For the teachers identified as ineffective based on three-year estimates, approximately half of them will continue to be identified as ineffective three years later.

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

PVAAS in Practice

In 2012, another state agency using a methodology similar to PVAAS reported to its legislature that less than 1% of teachers moved from the most effective designation to the least effective designation from one year to the next based on three-year teacher value-added measures.

In other words, in using a robust and reliable statistical approach, like PVAAS, for teacher estimates, Pennsylvanian educators and policymakers can build insightful policies customized to the teachers in their schools, 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] Bill and Melinda Gates Foundation (2013). Measures of Effective Teaching Project "Ensuring Fair and Reliable Measures of Effective Teaching," p. 13.