MEASURING TEACHER EFFECTIVENESS

WITH THE PENNSYLVANIA VALUE-ADDED ASSESSMENT SYSTEM

A Doctoral Thesis Presented to the

Faculty of the School of Education

Drexel University

by

Naomi Bowen

In Partial Fulfillment

of the

Requirements for the Degree

of

Doctor of Education in Leadership and Management

May 2017

Drexel UNIVERSITY

Abstract

Measuring Teacher Efficacy with the Pennsylvania Value-Added Assessment System

Naomi Bowen, Ed.D.

Drexel University, May 2017

Chairperson: Mary Jo Grdina

The purpose of this research was to determine if the Pennsylvania Value-Added Assessment System Average Growth Index (PVAAS AGI) scores, derived from standardized tests and calculated for Pennsylvania schools, provide a valid and reliable assessment of teacher effectiveness, as these scores are currently used to derive 15% of the annual effectiveness rating assigned to PVAAS eligible mathematics and English language arts teachers. The research also sought to determine if teacher perceptions indicate that the validity and reliability of the PVAAS AGI score included in the Classroom Rating Tool is important. The PVAAS AGI scores, average Normal Curve Equivalent (NCE) scores, and demographic data from 260 Pennsylvania middle schools serving grades six through eight were utilized to determine the extent to which PVAAS AGI scores assigned to a school change statistically over time and if there is a relationship between the PVAAS AGI in mathematics or English language arts for each school and the percentage of economically disadvantaged, learning disabled, English language learners, and minorities attending the school. A Likert scale survey of twenty PVAAS eligible teachers of mathematics and English language arts employed in two middle schools serving grades six through eight in the same district yielded information regarding teacher perceptions. Results of this study indicated that the mean gains for PVAAS AGI scores significantly decreased from 2013 to 2016 for all schools, regardless of achievement level, in both mathematics and English

language arts. The demographic variables analyzed did not impact PVAAS AGI scores for schools. Teacher morale was significantly negatively impacted by the inclusion of PVAAS AGI scores in the Classroom Rating Tool, the Pennsylvania mandated rubric utilized to measure teacher effectiveness. Additionally, teachers indicated that they believe important decisions such as changes to instructional practices and scheduling are made based on PVAAS AGI, even though teachers do not understand how it is calculated or trust the validity of PVAAS AGI scores. The literature review surfaced concerns regarding how the average NCE scores are treated in the PVAAS statistical model, as certain treatment of these scores could potentially force invalid results. The literature review also revealed that the Data Recognition Corporation (DRC), the company that designs the standardized tests from which the data utilized in the PVAAS statistical model is derived, has warned that scores at the maximum and the minimum end of the scale may not be accurate due to the design of their test. This is a matter of interest due to the fact that the SAS Institute claims that the tests utilized by their model must be able to measure the performance of students at the lowest and highest ends of the achievement spectrum, begging the question of whether it is even possible for the PVAAS model to accurately determine student growth. Finally, discrepancies in the reported grade levels for schools were found on the state data reporting site.

DREXEL UNIVERSITY
Graduate
College

DISSERTATION/THESIS APPROVAL FORM

This form is for use by all doctoral and master's students with a dissertation/thesis requirement. Please print clearly as the library will bind a copy of this form with each copy of the dissertation/thesis. All doctoral dissertations must conform to university format requirements, which is the responsibility of the student and supervising professor. Students should obtain a copy of the Thesis Manual located on the Graduate College or library website.

Dissertation/Thesis Title: Measuring Teacher Effectiveness with the Pennsylvania Value Added Assessment System

Author: Naomi Bowen

This dissertation/thesis is hereby accepted and approved.

Signatures:

Examining Committee

Chair

Members

Andrea Shaw Digitally signed by Andrea Shaw
DN: cn=Andrea Shaw, o=School of Education, ou=EdD Advisor, email=aes68@drexel.edu, c=US
Date: 2017.05.19 11:36:28 -04'00'

Academic Advisor

Department Head

Graduate College 3141 Chestnut Street Main 301 Philadelphia, PA 19104
Tel: 215.895.0366 Fax: 215.895.0495 Email: graduatecollege@drexel.edu Web: www.drexel.edu/graduatecollege

(Last Updated 3/1/2016)

Dedication

I dedicate this dissertation to my four beautiful children, Josephine Hunter, Savannah LeeAnn, Gabriel Alexander, and Adriana Marie, with the hope that my work inspires them to be life-long learners.

ACKNOWLEDGEMENTS

I am grateful for all of the guidance and support I received in designing and conducting my research study from Dr. Mary Jo Grdina, Dr. Allen Grant, Dr. Toni Sondergeld, and Dr. David Johnson. Thank you for your expertise and diligence in carefully reviewing my research, answering my questions, and always being open and willing to support my efforts. I was also very fortunate to share the company of Mary Ann Communale, Mindy O'Mealia, as we completed our doctoral journeys together. I truly enjoyed working together and having friends to turn to during this exciting adventure. I am thankful for women who have served as both mentors and close friends in education for many years. Kimberly Siar taught me how to break down blocks and look for creative solutions to help every child. Vivian Loewenstern entrusted me with my first leadership role and continued to support my endeavors long after our careers parted paths. Kathy Chesen's faith in my work, support, and willingness to lend an ear were integral to my success in this doctoral program and life.

It is truly an advantage to be born into a family that values life-long learning. My grandmother, Marianne Awerbuch, a survivor of the holocaust, returned to Berlin, Germany after raising my father in Israel, to finish her graduate work. She retired as the Dean of the Philosophy Department at the Free University of Berlin and impressed upon me two lessons: 1) There is no struggle that can't be overcome and 2) Never be afraid to speak out when a wrong must be made right. My parents, Jonathan and Patricia Awerbuch, professors of Engineering and Economics respectively, reinforced these lessons by modeling the importance of life-long learning, using their voices for the greater good, and supporting my academic endeavors. My sister, Dina Kasper, a lawyer, has helped me view the world through many different lenses.

This dissertation would not have been possible without the help and full support of my loving husband, Timothy, and my four beautiful children, Josephine Hunter, Savannah LeeAnn, Gabriel Alexander, and Adriana Marie. I hope to have passed down our family's value of life-long learning, as my children witnessed and supported my growth in academia. In turn, I am excited to enjoy and support the most important years of their lives, as Josephine and Savannah will soon be approaching college, discovering their identities and mission in this life. Thank you to my son, Gabriel, and my little girl, Adriana, for their patience as I worked to squeeze in play time between coursework, data analysis, and chapters; and for all of your love and support. I am thankful for my aspiring chef, Josephine, who always made sure I had dinner on the go, between chaperoning her and her siblings to their activities and completing my school work at night. I am thankful to Savannah for teaching me that a person can overcome even the greatest of difficulties with hard work and perseverance. Finally, thank you Timothy for your endless patience, love, and support.

Table of Contents

.

List of Tables

List of Figures

CHAPTER 1: Introduction to the Research

Historically, teacher effectiveness has been determined via classroom observation. In Pennsylvania, a rubric developed by Charlotte Danielson was adopted for administrators to utilize when rating observed teacher practices (Danielson, 2013). Over time, research studies have called into question the consistency and accuracy with which classroom observation rubrics can be utilized to measure teacher effectiveness. Recent research shows that classroom observation is a poor predictor of student performance (Cantrell & Kane, 2013). National policy changes over the past decade have resulted in shifting methodologies used for rating teacher effectiveness in Pennsylvania, as Pennsylvania attempted to attain consistency and accuracy in rating teacher effectiveness.

The passing of the No Child Left Behind Act (NCLB) in 2001 ignited a greater interest in student outcomes, causing states to recognize a need for ensuring that teachers produce positive student outcomes. The Pennsylvania Department of Education (PDE) implemented the Adequate Yearly Progress (AYP) system to measure the effectiveness of schools instructing student bodies, as well as specific subgroups, such as students in special education with individualized education plans (IEP), economically disadvantaged (ED), English as a second language learners (LEP), and minority groups (Pennsylvania Department of Education, 2012). In an attempt to meet the requirements of AYP, schools across Pennsylvania implemented data-driven instruction, a methodology through which student scores on the Pennsylvania State System of Assessment (PSSA) criterion referenced tests and the percentage of students scoring proficient or above for a given teacher, building, or school district were analyzed and used to make instructional decisions (Pennsylvania Department of Education, 2007). However, student performance did not yet factor into teacher effectiveness ratings and administrators continued to use classroom observation as the primary method for rating teacher effectiveness.

In 2012, the U.S. Department of Education began allowing states to request and receive waivers from the No Child Left Behind Act requirements under the condition, among others, that they integrate student achievement data into each teacher's effectiveness rating. In 2013, Pennsylvania joined the ranks of states approved for the ESEA waiver (Duncan, 2013). However, faced with the struggle to attain proficiency for 100% of its students, as had been required by NCLB, Pennsylvania adopted a new method for determining student achievement to comply with the ESEA waiver. This method used a statistical algorithm, also known as a value-added model (VAM), to measure growth from one standardized test to the next using normal curve equivalent scores derived from the PSSA, Pennsylvania's standardized criterion referenced reading and mathematics assessments. The Pennsylvania Value-Added Assessment System (PVAAS) became the new system through which the state of Pennsylvania utilized their VAM to calculate and report measures of student growth and to calculate the Average Growth Index, a measure of student growth across grade levels for a given teacher, building and district (Pennsylvania Department of Education, 2014).

The VAM utilized in the PVAAS system was purchased from the SAS Institute. The SAS Institute named the system utilizing their VAM the Education Value-Added Assessment System (EVAAS) (SAS Institute, 2014). The VAM utilized in EVAAS was originally developed by Dr. William Sanders, a professor who worked in the field of agricultural genetics at the University of Tennessee in the 1980s (University of Pennsylvania, 2004). When the EVAAS model was adopted by Pennsylvania it became known as the Pennsylvania Value-Added Assessment System (PVAAS). The PVAAS model specifically incorporates multiple years of student data to predict future normal curve equivalent scores on the state standardized tests by using normal curve equivalent scores from standardized tests in previous years. There are many types of VAMS, some taking into account student, teacher, or building variables (Wiley, 2006). Although possible, the PVAAS model does not take into account any variables, including demographics, developmental stages, English language proficiency, and class size (Commonwealth of Pennsylvania, 2015). The SAS Institute and Pennsylvania Department of Education argue that it is not necessary for PVAAS to take into account any variables because students are measured against their own normal curve equivalent score, thereby negating the need to account for variables (PDE Statewide Team, 2014). Independent researchers argue that this methodology does not negate the need to account for variables (Amrein-Beardsley, 2008). Therefore, a debate brews over the accuracy of utilizing the PVAAS system to measure teacher effectiveness in Pennsylvania.

Despite uncertainty regarding the accuracy of utilizing the PVAAS model, Pennsylvania has employed the model within its Classroom Rating Tool, a new system for measuring teacher effectiveness. On the Classroom Rating Tool, 50% of a teacher's score is derived from classroom observation and 15% is derived from a statistical method that involves calculating the change in students' growth scores of the given teacher's students. This is called the PVAAS Average Growth Index (AGI). Another 15% is derived from the PVAAS AGI for the entire school building and 20% is derived from elective data using Student Learning Objectives (SLO) which are developed and measured by each teacher (Research for Action, 2013). Figure 1 shows Section C of the Classroom Rating Tool utilized by the Pennsylvania Department of Education to measure teacher efficacy. The Building Level Rating in Section C is derived from the PVAAS AGI calculated for the building and comprises 15% of the teacher's overall rating and the

Teacher Level Rating in Section C is derived from the PVAAS AGI calculated for the teacher and also comprises 15% of the teacher's overall rating. Thirty percent of each teacher's rating is derived from a PVAAS AGI calculation. Although both are referred to as the PVAAS Average Growth index, it is important to note that the statistical algorithms utilized to calculate the Building Level Rating and Teacher Specific Rating are different.

(C) Final Teacher Effectiveness Rating – All Measures

Measure	Rating (C)	Factor (D)	Earned Points (C x D)	Max Points
(1) Teacher Observation & Practice Rating		50%		1.50
(2) Building Level Rating		15%		0.45
(3) Teacher Specific Rating		15%		0.45
(4) Elective Rating		20%		0.60
Total Earned Points				3.00

Conversion to Performance Rating	
Total Earned Points	**Rating**
0.00-0.49	Failing
0.50-1.49	Needs Improvement
1.50-2.49	Proficient
2.50-3.00	Distinguished
Performance Rating	

☐ **Rating: Professional Employee,** ***OR*** ☐ **Rating: Temporary Professional Employee**

I certify that the above-named employee for the period beginning ____________ (month/day/year) and ending ____________ (month/day/year) has received a performance rating of:

☐ **DISTINGUISHED** ☐ **PROFICIENT** ☐ **NEEDS IMPROVEMENT** ☐ **FAILING**

resulting in a FINAL rating of:

☐ **SATISFACTORY** ☐ **UNSATISFACTORY**

A performance rating of Distinguished, Proficient or Needs Improvement shall be considered satisfactory, except that the second Needs Improvement rating issued by the same employer within 10 years of the first final rating of Needs Improvement where the employee is in the same certification shall be considered unsatisfactory. A rating of Failing shall be considered unsatisfactory.

Date ____ Designated Rater / Position: ____ Date ____ Chief School Administrator ____

Figure 1: Portion of Pennsylvania Department of Education Classroom Rating Tool.

Statement of the Problem

Pennsylvania is currently utilizing a method for rating teacher effectiveness that may be inaccurate. The SAS Institute states that their EVAAS model, adopted by PDE and called PVAAS, to measure teacher effectiveness, does not need to account for variables such as student demographics, resources, and funding. National research has brought attention to the possibility that VAMs which do not account for variables in education may skew data (Amrein-Beardsley, 2008). If this is the case, an inaccurate measure of teacher effectiveness is being utilized to measure the efficacy of teachers in Pennsylvania. Furthermore, if schools are using inaccurate data to make staffing, scheduling, and programming decisions, this could result in negative outcomes for students.

The PVAAS model assumes that students should grow at a steady rate over time. However, it is now common knowledge that different students learn at different rates throughout different stages of their development and that growth does not progress at a linear rate (UNICEF, n.d.). Additionally, language acquisition occurs more quickly at younger ages. Younger English language learners may be able to acquire the English language more quickly than older English language learners (Phillips, 2002), giving them an advantage when growth scores are calculated. In Pennsylvania, school funding varies based on the number of economically disadvantaged students and special education students that a school district and school building serves.

In addition to potential variation in PVAAS AGI calculations as a direct result of the demographic factors, changes in funding as a result of these factors could also impact the PVAAS AGI. The data utilized in this study was obtained when schools were operating under the provisions of ESEA, which outlines in Title 1, the qualifications for receiving additional funding for economically disadvantaged students (U.S. Department of Education, 2015). States also factor the number of economically disadvantaged students and the number of students who speak English as a second language when calculating the amount of money dispersed across school districts (Basic Education Funding Commission, 2016). Other factors, such as median household income and the tax capacity index are also calculated into this formula, causing the amount of money allocated per student in each Pennsylvania school district to be different (Basic Education Funding Commission, 2016). Similarly, schools receive additional federal and state funding to provide services to students requiring special education (Pennsylvania Association of

School Business Officials, 2015). Schools are required to report the number of students considered economically disadvantaged and the number of students requiring special education on an annual basis (Pennsylvania Department of Education, 2016). These funding differences can potentially influence class size, the frequency of services, and the ratio of support personnel to students, as local school boards and administrators in Pennsylvania have some freedom with how they distribute funds to the students requiring support. Research has validated these concerns with the use of the PVAAS model including: a VAMs failure to account for class size, lack of random sampling of students, difficulty for students receiving high normal curve equivalent scores to make the same amount of growth as students with lower normal curve equivalent scores (also known as the ceiling effect), and the effects of resource allocation (Newton, Darling-Hammond, Haertel, & Thomas, 2010).

Purpose and Significance of the Problem

Purpose

The purpose of this research was to determine if the PVAAS AGI calculated for schools provides a valid and consistent assessment of teacher effectiveness and if teacher perceptions indicate that the validity and consistency of the PVAAS AGI score included in the Classroom Rating Tool to be important. In order to determine the accuracy and consistency of PVAAS AGI calculations this research will study whether the PVAAS AGI data produced is valid and reliable. “Validity refers to the degree to which something measures what it claims to measure. Reliability refers to the degree to which the measure is consistent when repeated (Harris, 2012, p. 3). Harris (2012) states that “determining validity and reliability is hard because there is no true measure to compare it to.” However, “A measure cannot be considered valid if it is heavily influenced by factors that are outside the control of teachers." (Harris, 2012). A large body of research on VAMs and a smaller body of research on the specific VAM utilized by PVAAS have yielded inconsistent results. Quantitative analysis of scoring data derived by PVAAS over a three-year span contributes to understanding the accuracy of this VAM in determining teacher effectiveness. Additionally, quantitative analysis of growth data and survey data determined if the utilization of the PVAAS system has contributed to changes in the school environment.

Significance

The advent of utilizing the PVAAS AGI in teachers' annual performance ratings has forced teachers of English language arts and mathematics to depend on the accuracy of their PVAAS AGI and their school building's PVAAS AGI in their subject area to receive a successful rating. If the VAM utilized by PVAAS does not produce accurate results or skews results for particular subgroups of students, the ratings that teachers receive will not accurately reflect teacher effectiveness. Although the SAS Institute refutes claims that their model may not provide accurate results, research on the accuracy of VAMs in education is gaining popularity (Commonwealth of Pennsylvania, 2015). Research on the relationship between classroom observation ratings and scores derived from a value-added model showed that there was a higher discrepancy between classroom observation ratings and value-added scores in schools that had higher rates of poverty. In this case, the value-added scores were significantly lower than the classroom observation ratings (Morgan, Hodge, Trepinksi, & Anderson, 2014). Another study focusing on whether disadvantaged students had less access to effective teachers, found that the teacher growth scores were lower when using a value-added model that did not account for student variables, indicating that disadvantaged students had less access to effective teachers. The researchers also noted that when using a value-added model that did account for student demographics, teacher scores were higher and it did not appear as though disadvantaged students had less access to effective teachers (Isenberg et al., 2013). The many variations of VAMs make it difficult to generalize research on different VAMs. In some cases, researchers develop their own VAM(s) which may or may not take into account student and building variables, such as demographics, developmental stages, language proficiency, and class size (Newton, Darling-Hammond, Haertel, & Thomas, 2010).

VAMs have not only been researched, but also used to support research on teacher efficacy across communities (Cantrell & Kane, 2013). However, in 2014, the American Statistical Association (ASA) released a statement generalized to all VAMs stating, "VAMs should always be accompanied by estimates of precision and a discussion of the assumptions and possible limitations of the model" (ASA, 2014, p. 1). The ASA also asserted that VAMs should not be used for individual teacher evaluation because they do not measure individual teacher contributions. VAMs do not measure what causes student growth, they can only show correlations. Different VAMs can also produce starkly different results. In light of conflicting research, Pennsylvania's implementation of a VAM in measuring teacher effectiveness raises many concerns in the education community.

Research Questions

This research will address concerns raised in the education community regarding the use of VAMs in rating teacher effectiveness by exploring the following questions:

1. To what extent do PVAAS AGI scores assigned to a school change statistically over time?

2. How does the PVAAS AGI scores assigned to a school relate to the percentage of economically disadvantaged, learning disabled, English language learners, and minorities in the school?
3. To what extent do teacher perceptions of the impact of PVAAS AGI indicate that the validity of the PVAAS AGI calculated for their school is important?

Conceptual Framework

A review of prior research and knowledge supporting the background of this study first focused on the characteristics of VAMS and previous research on types of VAMS, efficacy of different types of VAMs, and the pros and cons of using VAMS to measure teacher efficacy as indicated by previous research and professional organizations. Then, the research addresses policies and procedures leading up to the use of VAMs as a measure of teacher effectiveness. Finally, a review of student learning factors that could potentially influence student outcomes on assessments and their growth measures contributes to an understanding of why it may be necessary to account for student variables within a given VAM.

Researcher's Stance and Experiential Base

PDE's assertion that the PVAAS model implicitly accounts for student, building, and teacher variables begs several questions. First, if the model compares the students' growth from one year to the next, it seems impossible for the performance of the teacher in the previous year not to have an effect on the teacher's score the following year. For example, two students with the same teacher in a given year may produce entirely different gains depending on the efficacy of the teacher in the year before. The student with the ineffective teacher in the prior year may have achieved a lower percentile ranking, showing significant gains with the current teacher. On the other hand, the student who had a very effective teacher the year before may not show as much of a gain with the current teacher, because both teachers supported the student. Teachers have expressed a suspicion to this researcher that when students have lower scores in the year before, there is a higher likelihood that they will be able to achieve higher growth scores than when students have higher scores in the year prior. Anecdotal evidence from this researcher's growth scores as a 7th grade math teacher has supported this hypothesis. While pulling out the effects of different teachers is not within the scope of this research study, the experience has led the researcher to wonder if there are any other factors that are not related to an individual teacher's efficacy that may influence the PVAAS AGI received by the teacher.

Changing building initiatives can also shift areas of support, giving some teachers more support than others. For example, reading and mathematics specialists can be assigned to work with a specific grade level or class that has struggled on tests in the previous year. The students in this group may show more growth as a result of the extra support, benefiting the score of their teacher of record, and making it seem as if other teachers without the extra support were not as successful. This researcher served as a mathematics specialist for six years in an economically disadvantaged school district and saw firsthand how schools target areas of support with the goal of increasing test scores. Although providing the additional support is worthwhile to the students, it may influence the accuracy of a given teacher's PVAAS AGI. Similarly, English as a second language learners and economically disadvantaged students who received additional supports through federal and state funding may potentially experience growth that cannot be attributed to the teacher of record due to receiving extra services. Schools are complicated and ever evolving organizations that provide students learning support from many sources, yet the PVAAS AGI for teachers is assigned solely to the teacher of record.

Conceptual Framework of Research Streams

The literature review first takes a bird's eye view of existing research on value-added models and how some researchers have answered questions relating to the accuracy of using this type of model to measure teacher efficacy. The different types of value-added models will be reviewed to understand how differences in the type of model can impact outcomes. A review of research on teacher perceptions of the use of value-added models as a measure of teacher efficacy later informs conclusions about how the use of PVAAS AGI may impact teacher perceptions of changes in their school environment. Then, the research hones in on attributes of the specific model used in Pennsylvania. It reviews claims that the Pennsylvania Department of Education makes in supporting their use of the PVAAS model and identifies literature and evidence that supports or counters these claims.

A review of the policies leading up to the use of PVAAS in teacher evaluations will contribute to an understanding of the context. This illuminates how the importance of finding a valid means for measuring teacher effectiveness came to the forefront of education policy. The procedures developed as a result of these policies will provide insight into the impact and potential future impact of using this system of measurement in education.

Finally, a review of the factors that contribute to student learning provide important insight to the claims for and against student variables being accounted for in value- added models. Understanding all of the factors that contribute to student learning and how they change for a given student over time supports determining the validity of claims made by both sides of the debate. Figure 2 is a graphic representation of the conceptual framework utilized in this literature review.

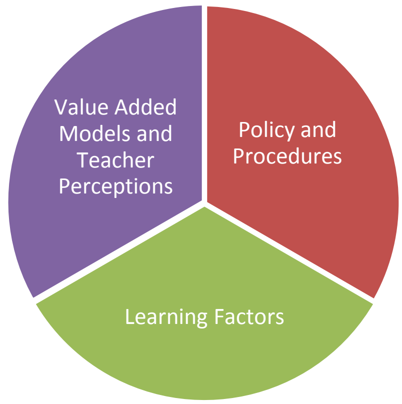

Figure 2: Conceptual Framework.

Definition of Terms

- Average Normal Curve Equivalent Score - The average of the Normal Curve Equivalent scores for every student in a school.
- Ceiling Effect- The decrease in ability to demonstrate growth as achievement increases.
- Criterion Referenced Assessment- An assessment aligned to specific content standards.
- Economically Disadvantaged Student (ED) - A student living at or below the poverty line for a given region.
- English as a Second Language Learner (ELL) - A limited English proficient student who was either not born in the United States or did not learn English as their first language.
- Normal Curve Equivalent Score - a score derived from the Pennsylvania System of School Assessment using each students scale score on the annual standardized test to rank each student on a normal curve.

- Norm Referenced Assessment - An assessment with vertically aligned content designed to compare a student's performance to the population.
- Pennsylvania System of Standardized Assessment - The system through which Pennsylvania develops and administers standardized assessments for English language arts, mathematics, and science.
- Pennsylvania Value Added Assessment System (PVAAS) - The system Pennsylvania uses to measure teacher, building, and district effectiveness.
- PSSA English Language Arts Achievement – The percentage of students in a school building that score proficient or advanced on the English Language Arts PSSA.
- PSSA Mathematics Achievement - The percentage of students in a school building that score proficient or advanced on the Mathematics PSSA.
- PVAAS Growth Measure - an estimate of a district's or school's influence on students' academic growth in each state assessed grade and subject area.
- PVAAS Average Growth Index (AGI) – A measure of student growth across tested grade levels in a district/school.
- Teacher Effectiveness - A measure of teacher performance based on the achievement of their students on a criterion-referenced assessment.
- Teacher of Record - The individual teacher responsible for providing the majority of instruction in a given subject area.
- Value-Added Model - Statistical measure used to determine the amount of growth a student demonstrates from one test to the next text.

Assumptions and Limitations

A great deal of variability exists across schools and school districts in Pennsylvania. Differences in demographics, resources, funding, and school structure are a few of the variables that need to be taken into account when attempting to isolate the appropriateness and impact of utilizing PVAAS scores in teacher evaluations. A significant deal of background knowledge is needed to understand the statistical model being discussed and evaluated in this research. In some cases the research relies on comparing explanations and descriptions of how these statistical models function by experts in the field of statistics. Additionally, this research relied on the accuracy of the data provided by PVAAS on their public website (SAS Institute Inc., 2016). This data was utilized with the assumption that PVAAS has reported all of the assessment data accurately, despite finding that some schools were categorized incorrectly by grade level. It is unknown whether the grade level categories reported on the website have any impact on the actual calculation of the PVAAS AGI scores provided on the website.

Summary

This research addressed questions surrounding the appropriateness of utilizing value-added models to measure teacher efficacy, with a specific focus on the model used in Pennsylvania, PVAAS. The validity of claims made by the Pennsylvania Department of Education regarding the accuracy of PVAAS scores was examined. This research also looked at teacher perceptions indicating the importance of the validity of PVAAS AGI scores being assigned to their school buildings.

Chapter 2: Literature Review

Recent federal and state policy changes have shifted the way in which teacher performance is determined. Prior to these changes, observation was the sole method of teacher evaluation. Today, a value-added model (VAM) measuring student academic growth is used to calculate 30% of mathematics or English language arts (ELA) teachers' effectiveness ratings. A VAM is a complex statistical algorithm that utilizes multiple years of individual student data from Pennsylvania's Standardized System of Assessment (PSSA) a series of criterion-referenced assessments in mathematics, English language arts, and science to determine if the student has demonstrated growth in their proficiency of the tested material. Therefore, teacher performance ratings in Pennsylvania now rely on the accuracy of this complex statistical algorithm, to ensure that their annual teacher effectiveness ratings are accurate.

There are many types of VAMs ranging from simple to complex. These VAMs may exclude or include different types of variables. Some VAMs take into account school effects such as class size and funding or covariates adjusting for student demographics while others do not. Pennsylvania purchased their system for calculating student growth with a VAM from the SAS Institute. The system distributed by the SAS Institute is called the Education Value Added Assessment System (EVAAS) and in adopting EVAAS, Pennsylvania renamed it the Pennsylvania Value Added Assessment System (PVAAS). The VAM utilized in PVAAS is called a layered model which excludes both school effects and covariates (McCaffrey, 2004). PVAAS and the Pennsylvania Department of Education (PDE) claim that it is not necessary for their VAM to include school effects and covariates because it is designed to provide intra-student correlation, thus negating the need to account for external factors. However, "The impact of omitted covariates on estimated teacher effects in the presence of intra-student correlation is subtle depending on both the distribution of the omitted covariates and the assignment of students to teachers" (McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004).

National research has brought into question the validity of utilizing VAMs to measure teacher effectiveness. Some research has shown that the VAM utilized by Pennsylvania skews the scores of teachers with high percentages of disadvantaged students, students with disabilities, or English language learners in their classrooms. Other research has shown that teachers with high percentages of students maintaining high scores on the criterion-reference tests are negatively impacted by a limiting factor called the ceiling effect. The purpose of this research is to determine if the value-added model utilized in Pennsylvania provides an accurate measure of teacher effectiveness.

The literature review is organized into three areas to support a broader understanding of the context, variables, and methods employed in measuring teacher effectiveness. The literature will first focus on the policies and procedures leading up to utilizing VAMs to measure teacher efficacy. Next, the types of VAMs, their strengths, weaknesses, and research on past utilization of VAMS in the context of education will be reviewed. Finally, factors that influence student learning and their subsequent performance on standardized assessments will be summarized to develop a greater understanding of whether there are any influences that may prevent a VAM that utilizes an intra-student correlation methodology to entirely negate the ability of outside influences impacting a score tied to teacher efficacy.

Changes in Pennsylvania Policy for Teacher Evaluation

The Education and Secondary Education Act (ESEA) of 1965 was part of Lyndon Johnson's "War on Poverty" (Sass, 2016). It provided federal funds to support the education of low-income students. This resulted in the development of educational programs such as Title 1, which provides supports to economically disadvantaged students and bilingual education (Sass, 2016). The ESEA was since reauthorized in 2001 with the No Child Left Behind Act (NCLB), in 2012 to include a flexibility waiver from NCLB, and again in 2015 with the Every Student Succeeds Act (ESSA). Each reauthorization of the original ESEA brought significant change to school and teacher accountability.

No Child Left Behind Act

The No Child Left Behind Act (NCLB) was passed in 2001, reauthorizing the ESEA of 1965. NCLB demanded greater accountability on the part of schools and teachers in response to growing concerns that the United States was being outperformed academically by countries around the world (Paige, 2002). NCLB authorized the government to allocate spending to improving American schools and tasked schools with the goal of leading every American child to proficiency in mathematics, English language arts, and science by the year 2014. Each state was required to set an expectation for yearly rate of growth. At that time rate of growth was defined as the percentage of total students scoring proficient or advanced on the mathematics, ELA, and science standardized exams. This measurement did not require measures of individual student growth. School districts were provided a report for the district as a whole and for each individual school detailing their progress (New America Foundation, 2014). Individual teacher efficacy was not linked to these measures; instead teacher efficacy was measured independently via classroom observation using the *Framework for Teaching* (Danielson, 2015).

Education and Secondary Education Act Waiver

The reauthorization of the Elementary and Secondary Education Act (ESEA) in 2012 provided states the opportunity to receive flexibility waivers from the No Child Left Behind Act requirements (United States Department of Education, 2012) under the condition, among others, that they integrate student achievement data into each teacher's performance rating. According to ESEA flexibility waiver requirements: "To receive this flexibility, an SEA and each LEA must commit to develop, adopt, pilot, and implement, with the involvement of teachers and principals, teacher and principal evaluation and support systems that: (1) will be used for continual improvement of instruction; (2) meaningfully differentiate performance using at least three performance levels; (3) use multiple valid measures in determining performance levels, including as a significant factor data on student growth for all students (including English Learners and students with disabilities), and other measures of professional practice (which may be gathered through multiple formats and sources, such as observations based on rigorous teacher performance standards, teacher portfolios, and student and parent surveys)…" (U.S. Department of Education, 2012). In 2013, Pennsylvania joined the ranks of states approved for the ESEA waiver (Duncan, 2013). The successful application for the waiver states that "data sources, including the Pennsylvania Value-Added Assessment System (PVAAS) will inform teacher effectiveness and related student achievement progress (Pennsylvania Department of Education, 2013, p. 25). The waiver also stated that "PVAAS teacher specific reporting will inform decisions about which teachers may function effectively in various roles (Pennsylvania Department of Education, 2013, p. 100). With the granting of Pennsylvania's flexibility waiver, PVAAS became the method through which Pennsylvania began using a VAM to measure individual student growth on assessments and tie this measure of growth to teacher and building efficacy ratings.

Race to the Top Awards

The American Recovery and Reinvestment Act (ARRA) was signed into law in 2009, allowing the federal government to invest in critical public sectors such as education. "The ARRA provided $4.35 billion for the Race to the Top (RTTT) fund, "a competitive grant program designed to encourage and reward states that are creating conditions for education innovation and reform; achieving significant improvement in student outcomes, including making substantial gains in student achievement, closing achievement gaps…." (U.S. Department of Education, 2009, p. 2). States were then requested to apply for grants from the RTTT fund which made several stipulations that married student assessment results and teacher evaluations. The RTTT application required states to prove that they would "Conduct annual evaluations of teachers and principals with data on student growth for their students, classes, and schools, …" (U.S Department of Education, 2009, p. 9). According to the RTTT guidelines: "Student growth means the change in student achievement (as defined in this notice) for an individual student between two or more points in time. A state may also include other measures that are rigorous and comparable across classrooms." (U.S. Department of Education, 2009, p. 14). The RTTT guidelines did not include any stipulations for ensuring the accuracy of a state's method for calculating growth. States wanting to dip into the RTTT fund were required not only to begin measuring individual student growth across assessments but to also begin tying these measures of growth, with no stipulations for ensuring the accuracy of the selected measure, to teacher evaluations.

Pennsylvania won their RTTT grant during the third phase of the RTTT grant cycle in 2011, receiving $41,226,299 (U.S. Department of Education, 2011). In order to comply with the grant requirements, Pennsylvania proposed that they would "refine and implement teacher and principal evaluation systems that incorporate student performance results as a significant factor" (U.S. Department of Education, 2011, p. 5). According to the application, Pennsylvania stated intentions to "tie PSSA and Keystone scores and PVAAS Growth Scores to individual teachers" (U.S. Department of Education, 2011, p. 41) and "establish clear approaches to measuring student growth and measure it for each individual student" (Race to the Top Application, 2011, p. 43). The application also stated that "Legislation is currently pending that would require 50% of an educator's evaluation be based on multiple measures of student achievement." Additionally, the RTTT grant application also described plans to develop a school report card based on an A-F grading system (U.S. Department of Education, 2011, p. 41). Winning the Race to the Top grant in 2011 launched a new age of teacher and school evaluation systems in Pennsylvania. However, despite following through with the terms of the grant, student proficiency rates increased in English language arts (ELA) but decreased in mathematics across the state (U.S. Department of Education, 2016).

Act 82

In compliance with the conditions of the ESEA flexibility waiver and the Race to the Top grant, Pennsylvania passed Act 82, amending the Pennsylvania Public School Code of 1949. The teacher effectiveness scoring guidelines outlined by Act 82 (PA Bulletin, 2013) were translated by the Pennsylvania Department of Education (PDE) into the Classroom Teacher Rating Tool (Pennsylvania Department of Education, 2015), a form on which to calculate a teacher's combined score from observation, student growth achievement, and elective data. On the Classroom Teacher Rating Tool (Appendix A), 15% of the performance rating for teachers of math, English language arts, and science is determined by the teacher's ability to help students demonstrate growth on standardized tests. Another 15% of a teacher's performance rating is derived from the ability of the entire building to demonstrate growth on standardized tests. Both scores are determined by PVAAS. Therefore, 30% of a teacher's performance rating is derived from a VAM calculation provided to each school by PVAAS.

Value-Added Models for Teacher Evaluation

Dr. William Sanders, a professor who worked in the field of agricultural genetics at the University of Tennessee developed a layered value-added model (VAM) in the 1980s and received approval to test this model using data from students in Knox County schools in Tennessee (University of Pennsylvania, 2004). In 1992, Sanders's VAM was adopted into Tennessee's Educational Improvement Act and was used across the state to measure student progress. The system for measurement was named the Tennessee Value-Added Assessment System (TVAAS) and would later become known as the Education Value-Added Assessment System (EVAAS) when it was adopted by the SAS Institute and the Pennsylvania Value-Added Assessment System (PVAAS) when implemented in Pennsylvania. For the purpose of this study, it is important to understand that TVAAS, EVAAS, and PVAAS all refer to the same system of measuring student growth using the VAM developed by William Sanders.

"TVAAS uses scale scores from the norm-referenced items on the Tennessee Comprehensive Assessment Program (TCAP)" (Sanders & Horn, 1994). The Tennessee Comprehensive Assessment Program integrated questions from the California Achievement Test (CTBS/4) and it was the results of answers to the questions from CTBS/4 that were utilized in the VAM produced by Sanders (Sanders & Horn, 1994). "The California Achievement Test is a nationally normed standardized test that measures achievement in the areas of reading, Language Arts, and math" (Seton Testing Services, 2016). Sanders found that TVAAS produced values for teacher effects that strongly correlated with subjective evaluation by supervisors and that these values for measuring school and teacher efficacy tended to be consistent (Sanders & Horn, 1994). Utilization of the value-added model to measure student achievement growth quickly spread across the nation as other states sought to meet the requirements for the ESEA flexibility waiver and Race to the Top. Today, the VAM developed by Sanders is marketed by the SAS Institute as the Education Value-Added Assessment System (EVAAS) (SAS Institute, 2014). Several states adopted this model and renamed it for their state. In Pennsylvania this model became known as the Pennsylvania Value- Added Assessment System (PVAAS) (SAS, 2016). However, the testing data utilized within the model has varied across states. While the research conducted in Tennessee utilized Sanders's VAM with the data from only Tennessee's norm-referenced assessment items, Pennsylvania utilizes Sanders's VAM with data from the Pennsylvania State System of Assessment (PSSA) which administers only criterion-referenced assessments. This fact is significant because Sanders's VAM is designed to measure growth from one year to the next on norm-referenced assessments which differ significantly from the criterion-referenced assessments utilized in some states. The purpose of a norm-referenced assessment is to "To rank each student with respect to the achievement of others in broad areas of knowledge" and "to discriminate between high and low achievers" (Huitt, 1996). A norm-referenced test "measures broad skill areas sampled from a variety of textbooks, syllabi, and the judgments of curriculum experts (Huitt, 1996). The purpose of criterion referenced assessments is "to determine whether each student has achieved specific skills or concepts" and "to find out how much students know before instruction begins and after it has finished (Huitt, 1996). A criterion referenced assessment "measures specific skills which make up a designated curriculum. These skills are identified by teachers and curriculum experts" and "each skill is expressed as an instructional objective" (Huitt, 1996). The criterion referenced assessments administered by Pennsylvania require students to demonstrate proficiency in very specific state standards-aligned skill sets which vary from year to year. Research on the utilization of Sanders's VAM utilizing data from Pennsylvania's criterion-referenced assessments could potentially produce different results than the research conducted by Sanders and Horn using data from Tennessee's norm-referenced assessment items in 1994. One of the issues that may arise with utilizing criterion-referenced assessment data as opposed to norm-referenced assessment data is described by Koedel and Betts (2010), "In practice it might be quite important whether a district uses a norm-referenced or a criterion-referenced test for the purpose of evaluating teaching effectiveness. A norm-referenced test is a standardized test that is meant to estimate where a student ranks against the test score distribution of the reference group, typically the national student population. Such a test, if well designed, should exhibit few ceiling effects because it must include questions with a range of difficulty so that distinctions can be made among students through-out the test score distribution....we speculate that these criterion-referenced tests are more likely to exhibit ceiling effects, particularly when a state exam is intended, either explicitly or implicitly, to serve as a minimum-competency test. For example, in Mississippi the state-level test appears to be aimed at a fairly low level. In 2006 to 2007, 90 percent of fourth-grade students scored at or above the "proficient" level in reading on the state-level Mississippi Curriculum Test (MCT). However, just 19 percent of these students scored at or above the proficient level on the National Assessment of Education Progress (NAEP)" (Koedel & Betts, 2009). Koedel and Betts (2009) concluded that "researchers and policy makers should be concerned when working in minimum-competency or proficiency-based testing environments. We show that ceiling conditions in such environments can significantly alter value-added assessments for individual teachers."

Defining Value-Added Models

A value-added model, also known as a VAM, is a statistical algorithm for determining and predicting growth in a data set over time. When applied to the education in Pennsylvania, a value-added model utilizes existing student data from criterion referenced assessments to determine annual growth and predict future test scores. The goal in using a value-added model is to determine factors associated with changes in student assessment data (Wiley, 2006).

There are many different variations of value-added models incorporating different pieces of data, causing debate over the type of model that should be used and which factors should be incorporated. The simplest of these models is the gain score model, a two score comparison, in which gains for all of a teacher's students are averaged. One limitation of this model is that it does not handle missing data, which can skew averages. The covariate adjustment model is similar to the gain score model, however incorporates an adjustment for student characteristics (Wiley, 2006).

The EVAAS model adopted by Pennsylvania is a layered model, also known as a multivariate model, incorporating multiple years of student data. This model is not as simple as the gain score model and covariate adjustment model because it requires multiple assessments to measure the effect of multiple teachers on student achievement over multiple years (Wiley, 2006). Although possible, the EVAAS model does not take into account variables such as demographics, developmental stages, English language proficiency, and class size (Commonwealth of Pennsylvania, 2015).

The American Statistical Association (ASA) endorses the use of value added models as a means for improving the quality of education within buildings and districts. However, the ASA states that, “VAMs should always be accompanied by estimates of precision and a discussion of the assumptions and possible limitations of the model” (ASA, 2014, page 1). The ASA also asserts that value-added models should not be used for individual teacher evaluation because it does not measure individual teacher contributions. Value-added models do not measure what causes student growth, they can only show correlations. Different value-added models can produce different results. Therefore, they are not useful for teacher evaluation purposes. The ASA states that studies have shown teachers as accounting for only 1-14% variability in test scores and that improvement of education is more likely to occur as a result of system level changes. For these reasons the ASA believes that value-added models can only be used for the purpose of quality improvement as opposed to assessment of teachers (ASA, 2014). According to Harris (2012), another factor that may influence the accuracy of value added models is that “value-added measures can also be biased, but in a somewhat different way. A common criticism of value-added measures is that some teachers are at a disadvantage because they are assigned students who are more difficult to education, even after the measures account for student’s prior test scores; this is what researchers call selections bias.”

Teacher Perceptions of Using Value-Added Models for Measuring Teacher Efficacy

VAMs involve complex statistical algorithms. One might wonder how many teachers have the background knowledge in statistics to understand the theory and methodology behind using VAMs to measure growth on criterion-referenced assessments. A study of teachers in the Southwest School District of Houston, Texas evaluated teacher perceptions of how the SAS EVAAS system works in practice for measuring student growth using the data derived from the standardized assessment administered in Texas. The Southwest School District is home to 300 schools, 204,000 students, and 11,000 teachers. Sixty two percent of the student are at risk, 92% are minorities, and 80% of the students are considered economically disadvantaged. The study was conducted using survey data collected from 882 teachers who were EVAAS eligible, meaning they were teachers of tested content areas and received EVAAS scores, in the Southwest School District (Collins, 2014, p.5). The study found that teachers experienced a fluctuation in their scores from one year to the next. "Among participants in this study, more teachers indicated that their SAS EVAAS® scores were inconsistent (n = 404/874; 46.2%) year-to-year than those who reported consistent scores (n = 371/874; 42.4%)" (Collins, 2014, p.9). This data damages the perceived reliability of EVAAS Average Growth Index (AGI) scores assigned to teachers. Some teachers reported that their scores varied based on the demographics of the students that they instructed: "Among the teachers who did provide an explanation for the fluctuation of their SAS EVAAS® scores, 24.4% (n = 93/381) reported the inconsistencies were caused by the different types of students they taught, and specifically referenced ELL and transition students as well as high achieving and gifted students as those responsible for score inconsistencies" (Collins, 2014, p.9). One ELL teacher reported: "Since I am teaching 5th grade ELL, I have been categorize[d] as ineffective because my students don't grow when coming from 4th grade all Spanish to 5th grade all English." (Collins, 2014, p. 9) A teacher of students classified as gifted reported: "The first year, they were ok. Then as I began to teach the gifted students, the scores continued to show negative growth. For the 2010-2011 school year, the Principal even told me that my scores revealed that I was one of the worst teachers in the school. The School Improvement Officer observed my teaching and reported that my teaching did not reflect the downward spiral in the scores" (Collins, 2014, p.9). This teacher's experience supports concerns of researchers that there exists a ceiling effect which may affect the validity of the EVAAS AGI assigned to teachers.

Some teachers had a positive perspective, attributing the positive change in their EVAAS AGI to their growth as a teacher: “My second year's score was higher than my first year's score. I attribute this to professional growth and experience.” (Collins, 2014, p. 10). Another teacher attributed the increase in their EVAAS AGI to learning how to teach to the test: “My first years of teaching I was still learning the ropes. Therefore, those scores were lower; however, over the years I understand that you must teach to the test to get the scores you want. To do well, the students must not only be intimate with the objectives, but also the lay-out and the verbiage on the test. Especially the ELL students. They need to know the wording of the questions beforehand so that they can be sure that they grasp what the question asks.” (Collins, 2014, p. 10).

A significant percentage of teachers also believed that demographics and home life seemed to influence their EVAAS AGI, “The final reliability question included in this section of the survey instrument asked teachers if they received consistent SAS EVAAS® scores despite the varied proportions of different types of students (i.e., ELL, gifted, special education, low/high income) they taught” (Collins, 2014, p. 10). Approximately 53% of teachers reported inconsistent EVAAS scores. Approximately 38% of teachers reported that student differences or external factors seemed to influence EVAAS scores (Collins, 2014, p.10). One teacher believed that: “[SAS EVAAS] depends a lot on home support, background knowledge, current family situation, lack of sleep, whether parents are at home, in jail, etc. There are too many outside factors – behavior issues, etc.” (Collins, 2014, p.10). Approximately 17% of teachers reported believing that gifted students had difficulty showing growth due to lack of room to grow on the test (Collins, 2014, p.10). Ultimately, the researchers concluded that utilization of EVAAS in the Southwest School resulted in a number of negative effects. Teachers reported the practice of gaming or using a variety of strategies to acquire the best rosters to get the highest EVAAS AGI. (Collins, 2014, p. 18). Other teachers reported that the use of EVAAS AGI resulted in a more competitive environment and lower morale among teachers (Collins, 2014, p.19). Teachers also expressed a lack of trust in the EVAAS methods of calculating AGI. One teacher reported the opinion that, "Ultimately, there are no stated metrics and as such I don't trust that the people who assign this number are using this in my or my school's best interest. To use the lingo, the current system is not transparent. That makes me more resistant to data [or] a system that has the potential to be very useful for testing." (Collins, 2014, p. 19). One math teacher in the study reported: "I don't completely believe in it or trust that the calculations are valid. And even if the whole EVAAS operation is mathematically sound, I'm still not sure if it is all that important" (Collins, 2014, p. 19). Another added, "Since I don't find the reports consistent with my instruction, effort and quality of practice, I don't trust EVAAS reports." (Collins, 2014, p. 19). The findings of this study demonstrate that a need for further investigation into the impact of using EVAAS AGI calculations to inform instructions and determine teacher effectiveness on students, teachers, and schools.

Learning Factors

An accurate value-added model should mediate any external factors so that no particular set of circumstances outside of a teachers control can positively or negatively influence the value-added measure. Research on the PISA assessment results for Germany and Spain in 2000 has shown that class size has a large effect on reading achievement, while socio-economic status and absenteeism have smaller effects (Kotte, Lietz, & Lopez, 2005). Therefore, a thorough understanding of the factors that may influence student achievement may facilitate determination of their influence on the Pennsylvania Value Added Assessment System (PVAAS) method for calculating the Average Growth Index (AGI), a score that is included in each mathematics and ELA teacher's yearly evaluation of teacher efficacy in Pennsylvania.

Student Learning Factors

Poverty has been a long standing concern related to academics and the driving force behind Lyndon Johnson's "War on Poverty" with the development and signing of the Education and Secondary Education Act of 1965 (Sass, 2016). "A growing body of evidence indicates that effects of poverty on physiologic and neurobiologic development are likely central to poverty-related gaps in academic achievement and the well-documented lifelong effects of poverty on physical and mental health" (Blair & Raver, 2016). Another study indicated that socioeconomic status (SES) influenced structural brain development: "Environmental differences associated with SES may influence aspects of structural brain development during childhood and adolescence" (Merz, He, Sowell, & Noble. 2016). It may be that increased funding to provide extra teachers through Title 1 is not enough to support consistent student growth among the economically disadvantaged population. Therefore, a teacher without appropriate support may have more difficulty demonstrating growth among this subgroup of students. A study involving 250 teachers and their combined 3,500 students across six high schools in the San Francisco Bay Area demonstrated that differences in context, such as student demographics and class size, impacted a teacher's score when using both value added models that account for student variables and value added models that do not account for student variables (Newton, Darling-Hammond, Haertel, & Thomas, 2010). In this study, researchers found that there was a negative correlation between student achievement scores and the percentage of disadvantaged students in a teacher's class. Isenberg et al.'s (2013) study on access to effective teaching utilized data gathered from teachers of grades four to eight across 29 school districts. The researchers determined that when teachers with high percentages of disadvantaged students in their classroom were rated using a value-added model that did not account for student variables, it appeared that disadvantaged students had less access to effective teachers. However, when rated using a value-added model that accounted for student variables there was no difference in access (Isenberg et al., 2013). These two studies promote the need for further research on score stability when a teacher instructs a class with a high percentage of disadvantaged students one year and a class with a low percentage of disadvantaged students the next year.

Another factor associated with student growth for all populations is the changes in the rate of brain development in children and adolescents over time. Unlike the expectation by PVASS of a linear rate of growth, "Different brain structures mature at different rates and follow different paths…" (Semrud-Clikeman, 2016). Semrud-Clikeman (2016) warns that "In each stage of development, it is important for teachers to understand the relationship between neurological development and learning. This understanding is particularly important when there is a mismatch between development and educational expectations. The mismatch may be due to brain maturational differences or it can be due to a developmental disability." Understanding that brain development does not occur at the same rate across age levels or across the span of an individual students' life begs the question of whether it is appropriate to place an expectation of linear growth on student academic achievement when using PVAAS AGI to measure teacher efficacy.

Similarly, the rate of language acquisition can vary across age levels among English language learners. A study of 2,700 English language learners from kindergarten through eighth grade indicated that: “Differences in reading and math achievement between ELLs and native English speakers varied based on the grade at which English proficiency is attained. Specifically, ELLs who were proficient in English by kindergarten entry kept pace with native English speakers in both reading and math initially and over time; ELLs who were proficient by first grade had modest gaps in reading and math achievement compared to native English speakers that closed narrowly or persisted over time; and ELLs who were not proficient by first grade had the largest initial gaps in reading and math achievement compared to native speakers but the gap narrowed over time in reading and grew over time in math. Among those whose home language is not English, acquiring English proficiency by kindergarten entry was associated with better cognitive and behavioral outcomes through eighth grade compared to taking longer to achieve proficiency” (Halle, Hair, Wandner, & McNamara, 2012). If the rate of language acquisition not only varies as students age but also impacts student achievement it is unclear whether PVAAS’s methodology of intra-student correlation truly provides an accurate representation of teacher effectiveness.

PVAAS does not account for student variables, such as economic status, learning ability, English language proficiency, class size, and funding. The Pennsylvania Department of Education (PDE) claims that it is not necessary for Pennsylvania Value Added Assessment System (PVAAS) to account for student variables, such as economic status, learning ability, English language proficiency, class size, and funding. According to PDE, every student should be able to show at least one year’s worth of growth on standardized tests (Commonwealth of Pennsylvania, 2015). To validate this claim, PDE’s report displays a scatterplot showing that some schools with high percentages of disadvantaged students evidenced large amounts of growth. What was not included in this report was contextual information, such as class size, school funding, grants, and other factors that may have increased a school's ability to support student achievement. Also not included, were multiple years of data demonstrating consistency in the results.

Building Level Factors

In 2007, the Brookings Institute published research findings on the effects of class size and school size on student achievement (Ready & Lee, 2007). Students in kindergarten and first grade from 1,000 public and private schools were randomly sampled and their achievement data was tracked for two years. Telephone interviews with parents and teachers of the children tracked were conducted to support the quantitative research. The researchers used hierarchical linear modeling with a three level growth curve framework. In explaining why they used this model to measure student achievement, the researchers criticized the use of value-added models, stating that these models take a traditional approach assuming variance remains steady over time. The results of the study indicated that being in a large class was a disadvantage and lowered achievement levels. Children in the smaller classes gained more over time. The researchers also found that students who spoke English as a second language made greater gains than native English speaking students (Ready & Lee, 2007).

Another study by Stout (2013) focused specifically on sixty-five high schools in Pennsylvania who received high PVAAS AGI scores in reading concluded that there were several characteristics common these schools. These high school included remediation programs, utilized data informed decision-making practices, and provided targeted professional development to teachers. The same study showed a positive correlation between the percentage of economically disadvantaged students, the number of students in the school, and the PVAAS

AGI scores in mathematics. This may indicate that there are factors outside of what happens in a given teacher's classroom that may influence the PVAAS AGI calculated for that teacher.

According to the National Bureau of Economic Research, school finance reforms have caused dramatic shifts in the achievement level and life outcomes for students living in economically disadvantaged communities. However, these funding increases have had little to no effect on students living in non-poor families (Jackson, Johnson, & Persico, 2014). One factor in the area of funding is resource allocation. A study driven using data collected over two years from 180 Georgia school districts, revealed that allocation of funds had a small but statistically significant effect on student achievement (James et al., 2011). This Connexions Project research found that spending in the area of improvements in instructional services, teacher salaries, and benefits showed a positive effect on student achievement. Spending in the areas of pupil services and technology had a negative effect on student achievement.

Teacher Performance Factors

A two-year study on 35 teachers in the State of Kentucky examined the relationship between job-embedded professional development, teacher efficacy, and student achievement in mathematics found a positive relationship (Althauser, 2010). Althauser (2010) found that the rigor, frequency, and applicability of professional development provided to teachers differed across districts because the availability of these options was determined by the School Board's funding priorities and district administrators. The professional development provided to a teacher by the school district appears to affect student achievement by enhancing teacher's knowledge and skills and improving classroom teaching strategies which, in turn, raises achievement. A survey of 3,250 teachers participating in 80 professional development programs in Australia concluded that professional development with a strong focus on content and student issues in learning provided the greatest effect on student outcomes (Ingvarson, Meiers, & Beavis, 2005). One review of professional development programs finds that some programs provided no statistically significant effect on student achievement and others showed a significant positive effect on achievement (Yoon et al., 2007). Another study by Isabel (2010) focused on the professional development provided to teacher in the Metropolitan Nashville Public Schools yield no correlation between high quality professional development and the Average Growth Index provided by the Tennessee Value Added Assessment System (TVAAS). This research suggested that there is a question of whether the professional development helped or if TVAAS was capable of reflecting the changes that occurred as a result of the professional development (Isabel, 2010).

Analysis of data collected from a large, Midwestern, economically impoverished school district, revealed a statistically significant difference in achievement between students with teachers having an average of five through nine years of teaching experience, and students whose teachers had over 13 years of teaching experience (Vanderhaar, Munoz, & Rodosky, 2007). Research utilizing five consecutive years of data from 132 teachers in South Carolina showed that teacher effectiveness scores determined by their value-added model were also unstable (Morgan, Hodge, Trepinksi, & Anderson, 2014). Teacher growth scores in both studies fluctuated from year-to-year and course-to-course. This research suggested that there needs to be a greater understanding of how and why a teacher's efficacy can change across years and classes.

Summary

Incorporation of PVAAS AGI into teacher efficacy ratings in the Pennsylvania public school system has increased the necessity for an assurance of accuracy in score calculations.

Research conducted by the William Sanders, the developer of the VAM used by PVAAS has demonstrated consistency and positive correlations between AGI scores and other measures of teacher efficacy. However, this research was conducted using norm-referenced assessment data, whereas the VAM in Pennsylvania is applied to criterion-referenced assessment data. Research is required to determine if a ceiling effect may compromise the accuracy of the PVAAS AGI when utilizing data from criterion-referenced assessments.

Although the SAS Institute claims that it is not necessary to account for student, building, or teacher variables, research has shown that all of these variables impact student achievement. Additionally, the linear expectation for growth applied by VAMs cannot be applied to language acquisition or learning in general as the rate of brain development changes across age levels and from student to student. PVAAS's use of intra-student correlation to negate the effects of poverty on students who are economically disadvantaged does not account for the fact that these students may have a slower rate of growth in general due to differences in brain development.

The use of PVAAS AGI within schools for decision making also requires further study. Research has indicated that some teachers have experienced a lack of consistency in their value-added scores or correlations between their students' demographics and scores. This research also showed that schools attempted to manipulate rosters to achieve higher growth scores. Investigation is needed on the positive and negative impacts of utilizing a VAM to measure teacher efficacy in the Pennsylvania public school system. Further research on PVAAS and correlations between the PVAAS AGI and variables that exist in the public school system and PVAAS AGI is needed to determine the validity and reliability of using PVAAS AGI to measure teacher efficacy.

Chapter 3: Research Methodology

The Pennsylvania Department of Education uses a system called the Pennsylvania Value-Added Assessment System (PVAAS) to determine student growth from one standardized test to the next (SAS Institute, 2014). This measure of growth for individual students is called the PVAAS Growth Score. The PVAAS Growth Score is calculated using a statistical algorithm called a Value-Added Model (VAM) and the SAS Institute supplies the particular model used by the state of Pennsylvania in PVAAS (PVAAS Statewide Team for PDE, 2015). The model is very complex, however in the simplest of terms, it converts each student's test scores from the Pennsylvania State Standardized Assessments (PSSA) in English language arts, mathematics, and science into normal curve equivalent scores and then compares each student's normal curve equivalent score from one year to the next (PVAAS Statewide Team for PDE, 2015). PVAAS claims to utilize a variation of common statistical models such as the linear mixed model within their multivariate response model (Rivers, J.C., Sanders, W.L., Wright, J.T., & White, S.P., 2010, p. 4).

After obtaining the growth score for each student, PVAAS utilizes another model to determine a measure of growth for each teacher of English language arts and mathematics, as well as for each school building as a whole. This measure of growth is calculated for both teachers and school buildings using another complex statistical algorithm that in simple terms averages the growth of all of the students in a teacher's class(s) and all of the students in a school building. This measure of growth is called an Average Growth Index (AGI) (PVAAS Statewide Team for PDE, 2015).

As a result of the passing of Act 82 in 2012, the Classroom Rating Tool was mandated by the Pennsylvania Department of Education in the 2013-2014 school years for the purpose of measuring and reporting teacher effectiveness (PA Bulletin, 2013). The Classroom Rating Tool derives fifteen percent of each mathematics and English language arts teacher's effectiveness rating from the PVAAS AGI in mathematics or English language arts for all of the teacher's students and fifteen percent of a mathematics and English language arts teacher's rating from the PVAAS AGI in mathematics or English language arts for all of the students in the building (Pennsylvania Public School Code, 2012). As a result, teachers of mathematics and English language arts now rely on the PVAAS AGI calculated for their building to reflect an accurate measure of teacher effectiveness.

Determining whether a value-added model can provide an accurate measure of school and teacher efficacy has become a point of controversy among education researchers. There are many variables in education that may not be accounted for in any particular model's calculations. The Pennsylvania Department of Education, in conjunction with the SAS Institute, has released statements regarding their choice not to account for student variables within the PVAAS model (EVAAS, 2015). They also dispute that there is a need to account for these variables with the argument that the PVAAS model compares a student's performance to their own previous performance from one year to the next utilizing their normal curve equivalent score. They claim that since the student is being measured against their own standard there is no need to control for variables (EVAAS, 2015). However, inherent to the use of normal curve equivalent scores is the comparison of student performance across varying demographics and school environments in the state of Pennsylvania. Thus, if unaccounted for, it may be possible that any change to these factors for a given student, such as a sudden change in financial status or improving English proficiency could impact their PVAAS AGI.

The Pennsylvania State Education Association (PSEA) released a report detailing a number of concerns related to utilizing a value-added model to determine teacher efficacy including (Zwerling, 2012). One of the first concerns is the fact that a value-added measurement does not provide a complete account of teacher effectiveness. “Teaching has always been, and remains, both an art and a science” (Zwerling, 2012, p. 1, para. 5). PSEA is also concerned that the score calculated using a value-added model is only as valid as the assessments scores being utilized in the model, however the value-added model itself only calculates teacher effects on growth “…they cannot measure growth with respect to Pennsylvania Academic Standards” (Zwerling, 2012, p. 2, para. 1). The assessment data being utilized in Pennsylvania’s value-added model is not designed to measure growth from one year to the next, as with vertically aligned nationally norm-referenced tests. Instead, the data utilized in the PVAAS model is derived from the PSSA, a criterion referenced assessment that measures proficiency in discrete topics that are not vertically aligned across grade levels (DRC 2010, p.232, para. 3).

While the Data Recognition Corporation (DRC), the company that designs the PSSA, warns that scores at the maximum and the minimum end of the scale may not be accurate (DRC 2010, p. 232, para. 2) the SAS Institute claims that the tests utilized by their model must “…adequately measure the performance of both very low and very high achieving students” (Rivers, J.C., Sanders, W.L., Wright, J.T., & White, S.P., 2010, p.2). This begs the overarching question of whether it is even possible for the calculations produced by PVAAS to be accurate. The Pennsylvania Department of Education and PVAAS also use infographics to dispute the ‘ceiling effect’. The ‘ceiling effect’ refers to the phenomenon of high achievers struggling to make high growth as a result of decreasing room to make growth (EVAAS, 2015). They use scatter plots to demonstrate that schools with high numbers of high achievers are capable of making growth and that schools with high numbers of low achievers are also able to make growth as calculated by PVAAS (Commonwealth of Pennsylvania, 2015). However, these scatterplots do not depict how a school's value-added score school changes over time. They only show that a school can reach high growth and high achievement in a given year. They do not show whether a school can consistently demonstrate high growth and high achievement over time or whether a school can demonstrate high achievement and high growth two or more years in a row. Similar scatterplots are utilized to demonstrate that in any given year a certain percentage of economically disadvantaged students and other subgroups of students can demonstrate high growth and high achievement.

The purpose of this research was to determine the validity and impact of utilizing PVAAS growth scores for schools in the classroom rating tool to determine teacher effectiveness. The first phase of testing identified the consistency of PVAAS AGI scores received annually by school schools serving grades six through eight over four years. The second phase of testing determined if the demographic variables such as the percentage of economically disadvantaged students, the percentage of English as a second language students, the percentage of learning disabled students, or the percentage of minority students correlate to certain PSSA ACH or PVAAS AGI. Finally, in the third phase of testing, teacher perceptions of PVAAS AGI that indicate the importance of its validity was determined using a Likert scale survey.

The research attempted to answer the following questions:

1. To what extent do PVAAS AGI scores assigned to a school change statistically over time?
2. How does the PVAAS AGI scores assigned to a school relate to the percentage of economically disadvantaged, learning disabled, English language learners, and minorities in the school?
3. To what extent do teacher perceptions of the impact of PVAAS AGI indicate that the validity of the PVAAS AGI calculated for their school is important?

This chapter provides an overview of this quantitative research design and the rationale for the design. It provides a description of the site and population addressed in this research. A detailed description of research methods and ethical considerations is also provided.

Research Design and Rationale

The first phase of this quantitative explanatory correlational research design utilized PVAAS AGI data for Pennsylvania middle schools serving grades six to eight to determine the level of consistency in PVAAS AGI data assigned to each school building over a period of four years. The data utilized in this study was calculated by PVAAS, published on the PVAAS website (SAS Institute Inc., 2016), and available for public use. A mixed ANOVA was used to determine the consistency of the PVAAS AGI scores for schools at different achievement levels over a four year period.

The second phase of the research addressed the relationship between demographic factors and PVAAS AGI. A Pearson correlation test between each of the four identified subgroups and the PVAAS AGI for each middle school was conducted to determine if the value-added model truly provides a representation of student growth that is not influenced by the four demographic factors identified in this research study, as claimed by PDE and the SAS Institute (EVAAS, 2015). The following sets of data were compared:

1. Percent of students categorized as economically disadvantaged per building vs. the building's PVAAS AGI in English language arts and mathematics.
2. Percent of students identified as having a specific learning disability vs. the building's PVAAS AGI in English language arts and mathematics.
3. Percent of students identified as speaking English as a second language vs. the building's PVAAS AGI in English language arts and mathematics.
4. Percent of student identified as minority vs. the building's PVAAS AGI in English language arts and mathematics.

The underlying assumption was that if the PVAAS AGI provides an accurate assessment of student learning, the Pearson correlation test would demonstrate no correlation between any of these demographic subgroups and student growth scores. If the research illuminates correlations between PVAAS AGI and a given demographic variable, then it can be understood that the PVAAS AGI does not provide a pure representation of student growth that excludes factors not associated with the specific student's achievement.

The third phase of the analysis looked at whether teachers perceive the importance of having valid and consistent PVAAS AGI score data included in annual evaluations. Data from a survey utilizing a Likert scale design was analyzed to determine if there was any concern on the part of teachers related to the validity and consistency of PVAAS AGI data.

Site and Population

The data collected and analyzed in the first and second phase of this research study was derived from the PSSA mathematics and English language arts scores of students in grades six through eight in the state of Pennsylvania. All of the data used to determine trends in variance and to identify consistency was published on the PVAAS website for public use (SAS Institute, 2016). The data utilized in this study included demographic data for each school, average NCE scores, and PVAAS AGI data in mathematics and English language arts for students in public middle schools that serve only six through eighth grade.

The purpose of selecting this particular population is to negate the effects of changes attributed to differing grade level tests calculated into the PVAAS AGI, age of transition to middle school, and other factors that may differ across schools that serve different age and grade level ranges. Therefore, public middle schools that serve grades five through eight or serve grades seven through nine were not included in this study.

Only mathematics and English language arts data were included in this study because students receive the mathematics and English language arts PSSA test in grades six, seven, and eight, providing three years of data produced at regular intervals. The science PSSA is not administered on a yearly basis resulting in a difference in the way that the PVAAS AGI in science is calculated for middle schools (Pennsylvania Department of Education, 2016).

Elementary schools were not included in this study because the first PSSA test provided is in third grade and there are not enough years of data to obtain optimal growth scores on which to base the PVAAS AGI, as the PVAAS AGI calculation requires three years of data from the same students (PVAAS Statewide Team for PDE, 2015). High schools were not included in this study because they do not administer the PSSA test in mathematics and English language arts on a yearly basis (Pennsylvania Department of Education, 2016). Gaps in years of testing could potentially influence the outcome of the analysis. Private schools, alternative schools, and intermediate units were not included in this analysis to negate the influence of systemic differences across these different types of school systems. Charter schools were not included in this study.

Middle schools serving grades six through eight were chosen to be the focus of this study due to the fact that the majority of the students whose test scores were included in the analysis have at least three years of prior data from grades three through seven included in their PVAAS growth score calculation which is in turn utilized in the building's PVAAS AGI calculation as the first PSSA test is not provided to students until third grade (Pennsylvania Department of Education, 2016). This was necessary to ensure that the method utilized to calculate PVAAS AGI was the same for all data included, as slightly different models are used to calculate science growth scores, Keystone Algebra growth scores and the growth scores for other standardized tests (PVAAS Statewide Team for PDE, 2015). Sixth grade students and students at grade levels above sixth grade have had the opportunity to take at least three PSSA assessments in third, fourth, and fifth grades. There were no issues with site access for this phase of data collection because all of the data utilized is currently published for public use by PVAAS (SAS Institute, 2016).

The third phase of the research utilized data collected from the Likert scale survey. This survey data was gathered from middle school math and English language arts teachers in sixth, seventh, and eighth grade at two middle schools in the same school district. The survey was provided to teachers via Google forms.

Research Methods

In this study, a mixed ANOVA was utilized to determine the consistency of PVAAS AGI data across multiple years. A Pearson Correlation test was utilized to determine if there were any correlations between the percentage of students in each of the four identified demographic subgroups per school building and the PVAAS AGI score for the associated school building. Finally, descriptive statistics was utilized to evaluate the data collected from the demographic questionnaire and Likert scale survey provided to PVAAS eligible teachers of mathematics and English language arts in grades six through eight.

Phase 1: Consistency in Average Growth Index Scores

A mixed ANOVA design was used to measure the stability of PVAAS AGI over a four year period for each school included in this research study. The mixed ANOVA was utilized to identify possible differences between related means in the case where the dependent variable is continuous and the independent variable is categorical (Lund, 2013). Two independent mixed ANOVA tests will be performed to assess the stability of annual PVAAS AGI scores in mathematics and English language arts from 2013 to 2016 for the 260 Pennsylvania public schools serving grades six to eight. In order to identify consistency of PVAAS AGI in mathematics and English language arts, the schools will be treated as the dependent variable and categorized based on their average Normal Curve Equivalent (NCE) scores in mathematics or English language arts using data from the baseline year, 2013.

Data collection. The first phase of the research had one stage of data collection, as all of the data utilized was currently available for public use on the PVAAS website (SAS Institute, 2016). PVAAS provides scatterplots comparing average Normal Curve Equivalent scores and PVAAS

AGI data for each school in the state of Pennsylvania from 2013 through 2016 on their public website. (SAS Institute, 2016). The tool used to create the scatterplots also permits the user to create a table of the data.

PVAAS also provides demographic information for each school in the current year, average Normal Curve Equivalent (NCE) scores for each school, PSSA achievement scores, and PVAAS AGI scores for the 2013 through 2016 school years for each school in the state of Pennsylvania. The data on the website is organized to display school name, school district, grades served, percentage of student proficient on the PSSA, and PVAAS AGI scores in table columns that can be sorted to identify only schools that serve grades six through eight. The average NCE scores for each school building were acquired from the scatter plot tool located on the website.

Data analysis procedures. In the mixed ANOVA tests identifying stability in PVAAS AGI scores in English language arts, the schools were categorized by achievement level. A school was categorized as “ELA Level 1” if the average Normal Curve Equivalent (NCE) score for all students in the school building was greater than 60% in the baseline year of 2013. Schools were categorized as “ELA Level 2” if the average NCE score on the PSSA in English language arts was greater than or equal to 50% and less than 60%. Schools were categorized as “ELA Level 3” if the average NCE score on the PSSA in English language arts was greater than or equal to 40% and less than 50%. Schools were categorized as “ELA Level 4” if the average NCE score on the PSSA in English language arts was greater than 30% and less than 40%. Finally, schools were categorized as “ELA Level 5” if the average NCE score on the PSSA in English language arts was great than or equal to 20% and less than 30%. None of the schools analyzed had an average NCE less than 20% or greater than 70% in English language arts.

Similarly, in the repeated measures ANOVA tests identifying stability in PVAAS AGI for mathematics, schools were categorized by achievement level using the same constraints. Schools were be categorized as “MATH Level 1” if the average NCE score in mathematics was greater than or equal to 60% in the baseline year of 2013. Schools were categorized as “MATH Level 2” if the average NCE score on the PSSA in mathematics was greater than or equal to 50% and less than 60%. Schools were categorized as “MATH Level 3” if the average NCE score on the PSSA in mathematics was greater than or equal to 40% and less than 50%. Schools were categorized as “MATH Level 4” if the average NCE score on the PSSA in mathematics was greater than 30% and less than 40%. Finally, schools were categorized as “MATH Level 5” if the percentage of students scoring proficient on the PSSA in mathematics was greater than 20% and less than 30%. None of the schools analyzed had an average NCE less than 20% or greater than 70% in English language arts. Table 1 and Table 2 provide an overview of the methodology used for categorizing schools in each test.

Table 1

Method for categorizing schools based on PSSA performance in English language arts

English Language Arts Level	Achievement in English Language Arts
1	Average NCE in English language arts ≥ 60
2	50 ≤ Average NCE in English language arts < 60
3	40 ≤ Average NCE in English language arts < 50
4	30 ≤ Average NCE in English language arts < 40
5	20 ≤ Average NCE in English language arts < 30

Table 2

Method for categorizing schools based on PSSA performance in mathematics

Mathematics Level	Achievement in Mathematics
1	Average NCE in Mathematics ≥ 60
2	50 ≤ Average NCE in Mathematics < 60
3	40 ≤ Average NCE in Mathematics < 50
4	30 ≤ Average NCE in Mathematics < 40
5	20 ≤ Average NCE in English language arts < 30

The mixed ANOVA test utilized to determine consistency is more specifically referred to as a 5 between x 4 within repeated measures ANOVA. This is due to the fact that five achievement levels are being assessed for stability of their growth measures across four annual growth measures from the year 2013 to 2016.

Null Hypotheses

$H_0$1 There will be no difference in the mean gain score of schools categorized as ELA Level 5, ELA Level 4, ELA Level 3, ELA Level 2, and ELA Level1.

$$H_01_: \mu_{\text{ELA Level 5}} = \mu_{\text{ELA Level 4}} = \mu_{\text{ELA LEVEL 3}} = \mu_{\text{ELA Level 3}} = \mu_{\text{ELA LEVEL 2}} = \mu_{\text{ELA Level 1}}$$

H02 There will be no difference in the mean gain scores of schools categorized as Math Level 5, Math Level 4, Math Level 3, Math Level 2, and Math Level 1.

H_02: $\mu_{\text{MATH Level 5}} = \mu_{\text{MATH LEVEL 4}} = \mu_{\text{MATH LEVEL 3}} = \mu_{\text{MATH LEVEL 3}} = \mu_{\text{MATH LEVEL 2}} = \mu_{\text{MAH LEVEL 1}}$

The Statistical Package for Social Sciences (SPSS) was utilized to conduct mixed ANOVA tests and yielded partial eta square (η_p^2) effect sizes. Effect size was used to determine the extent of the relationship between the achievement level and the change in PVAAS AGI over the four year period for each test. The significance level for each test was set at 0.05%. Effect sizes were interpreted as follows: small, $0.01 \leq \eta_p^2 < 0.06$; medium, $0.06 \leq \eta_p^2 < 0.14$; large, $0.06 \leq \eta_p^2 < 0.1$. Table 3 provides an overview of the tests utilized to determine consistency of PVAAS AGI scores assigned to schools with differing levels of achievement in Pennsylvania.

Table 3

Summary of tests determining consistency of PVAAS AGI scores

Test	Independent Variable	Dependent Variable	Treatment
1	English language arts achievement level	PVAAS AGI in English language arts	Mixed ANOVA
2	Mathematics achievement level	PVAAS AGI in Mathematics	Mixed ANOVA

Phase 2: Relationships between Demographics and the PVAAS AGI Scores

Pearson correlation tests were used as the primary method to determine the relationships between PVAAS AGI in English language arts and mathematics and each of the demographic subgroups. The Pearson correlation test was used to measure and describe the relationship between two continuous variables. In each of the tests, the independent variable was the percentage of students in a building that belong to a given demographic subgroup and the dependent variable was the PVAAS AGI score for the building. Both the percentages of students in a given demographic subgroup and the PVAAS AGI scores for buildings provide data that is continuous. Since PVAAS AGI is calculated using NCE scores, prior to performing the primary set of Pearson correlation tests designed to directly answer the research question regarding correlations between PVAAS AGI and demographics, a secondary set of Pearson correlation tests was performed comparing the average NCE scores of each school and demographic variables. The purpose of these tests was to acquire additional information that may place the results of the primary tests in context.

Data collection. The data utilized is this second phase of the research was copied directly from the data tables provided by PVAAS and the demographic data that was collected from PVAAS in the first phase of research. The data was compiled into Excel spreadsheets. The first Excel spreadsheet contained the following information: school name, grades tested, demographic information from the 2015 school year, average NCE scores in mathematics, PSSA achievement score data for mathematics, and PVAAS AGI data for mathematics from the 2015 school year. The second Excel spreadsheet contained the following information: school name, grades tested, demographic information for the 2015 school year, average NCE scores in English language arts, PSSA achievement score data for English language arts, and PVAAS AGI data for English language arts for the 2015 school year. The third Excel spreadsheet contained the following information: school name, grades tested, demographic information from the 2016 school year, average NCE scores in mathematics, PSSA achievement score data for mathematics, and PVAAS AGI data for mathematics from the 2016 school year. The fourth Excel spreadsheet contained the following information: school name, grades tested, demographic information for the 2016 school year, average NCE scores in English language arts, PSSA achievement score data for English language arts, and PVAAS AGI data for English language arts for the 2016 school year.

Data analysis procedures. The data in each of these Excel files was then used to run each of the Pearson correlations tests. In a Pearson correlation test the degree of the relationship is defined by the correlation coefficient, denoted r, and falls between the values of -1 and 1. If $r = 0$ no relationship between the two variables is indicated. If $r = +1$, there is a perfectly positive relationship between the two variables and if $r = -1$, there is a perfectly negative relationship between the two variables. The following guidelines can be used to interpret the r statistic in terms of the value of the relationship: very weak, $.00 \leq |r| \leq 0.29$; weak, $.30 \leq |r| \leq 0.49$; moderate, $.50 \leq |r| \leq 0.69$, strong; $.70 \leq |r| \leq 0.89$; very strong, $.90 \leq |r| \leq 1.0$.

Null Hypotheses

$H_0 1$: No correlation exists between the percentage of economically disadvantaged students in a school and the school's PVAAS AGI in mathematics.

$H_0 1$: $r_{ed/math} = 0$

$H_0 2$: No correlation exists between the percentage of economically disadvantaged students in a school and the school's PVAAS AGI in English language arts.

$H_0 3$: $r_{ed/ela} = 0$

$H_0 3$: No correlation exists between the percentage of students with learning disabilities in a school and the school's PVAAS AGI in mathematics.

$H_0 3$: $r_{iep/math} = 0$

$H_0$4: No correlation exists between the percentage of students with learning disabilities in a school and the school's PVAAS AGI in English language arts.

$H_04: r_{iep/ela} = 0$

$H_0$5: No correlation exists between the percentage of students that are minorities in a school and the school's PVAAS AGI in mathematics.

$H_05: r_{min/math} = 0$

$H_0$6: No correlation exists between the percentage of students that are minorities in a school and the school's PVAAS AGI in English language arts.

$H_06: r_{min/ela} = 0$

$H_0$7: No correlation exists between the percentage of students that are English as a second language learners in a school and the school's PVAAS AGI in mathematics.

$H_07: r_{esl/math} = 0$

$H_0$8: No correlation exists between the percentage of students that are English as a second language learners in a school and the school's PVAAS AGI in English language arts.

$H_08: r_{esl/ela} = 0$

A series of Pearson correlation tests were performed to test each of these hypotheses in 2015 and 2016. Table 4 represents the individual tests performed using data from 2015. Table 5 represents the individual tests performed using data from 2016.

Table 4

Summary of tests for 2015 school year performed in phase 2

Test	Data	Treatment
1	Average NCE score in Mathematics vs. Percent Economically Disadvantaged	Pearson Correlation
2	Average NCE score in English language arts vs. Percent Economically Disadvantaged	Pearson Correlation
3	Average NCE score in Mathematics vs. Percent Minority	Pearson Correlation
4	Average NCE score in English language arts vs. Percent Minority	Pearson Correlation
5	Average NCE score in Mathematics vs. Percent Limited English Proficiency	Pearson Correlation
6	Average NCE score in English language arts vs. Percent Limited English Proficiency	Pearson Correlation
7	Average NCE score in Mathematics vs. Percent Special Education	Pearson Correlation
8	Average NCE score in English language arts vs. Percent Special Education	Pearson Correlation
9	PVAAS AGI in Mathematics vs. Percent Economically Disadvantaged	Pearson Correlation
10	PVAAS AGI in English language arts vs. Percent Economically	Pearson Correlation

	Disadvantaged	
11	PVAAS AGI in Mathematics vs. Percent Minority	Pearson Correlation
12	PVAAS AGI in English language arts vs. Percent Minority	Pearson Correlation
13	PVAAS AGI in Mathematics vs. Percent Limited English Proficiency	Pearson Correlation
14	PVAAS AGI in English language arts vs. Percent Limited English Proficiency	Pearson Correlation
15	PVAAS AGI in Mathematics vs. Percent Special Education	Pearson Correlation
16	PVAAS AGI in English language arts vs. Percent Special Education	Pearson Correlation

Table 5

Summary of tests for the 2016 school year performed in phase 2

Test	Data	Treatment
1	Average NCE score in Mathematics vs. Percent Economically Disadvantaged	Pearson Correlation
2	Average NCE score in English language arts vs. Percent Economically Disadvantaged	Pearson Correlation
3	Average NCE score in Mathematics vs. Percent Minority	Pearson Correlation

4	Average NCE score in English language arts vs. Percent Minority	Pearson Correlation
5	Average NCE score in Mathematics vs. Percent Limited English Proficiency	Pearson Correlation
6	Average NCE score in English language arts vs. Percent Limited English Proficiency	Pearson Correlation
7	Average NCE score in Mathematics vs. Percent Special Education	Pearson Correlation
8	Average NCE score in English language arts vs. Percent Special Education	Pearson Correlation
9	PVAAS AGI in Mathematics vs. Percent Economically Disadvantaged	Pearson Correlation
10	PVAAS AGI in English language arts vs. Percent Economically Disadvantaged	Pearson Correlation
11	PVAAS AGI in Mathematics vs. Percent Minority	Pearson Correlation
12	PVAAS AGI in English language arts vs. Percent Minority	Pearson Correlation
13	PVAAS AGI in Mathematics vs. Percent Limited English Proficiency	Pearson Correlation
14	PVAAS AGI in English language arts vs. Percent Limited English Proficiency	Pearson Correlation

15	PVAAS AGI in Mathematics vs. Percent Special Education	Pearson Correlation
16	PVAAS AGI in English language arts vs. Percent Special Education	Pearson Correlation

Phase 3: Teacher Perceptions of Change Resulting from PVAAS AGI Scores

The third phase of this research study measured teacher perceptions indicating the importance of the validity and consistency of PVAAS AGI scores assigned to each school building. The following hypotheses were tested to answer the third research question.

Research Hypothesis 1: Teachers believe that there is a shift in school morale as a result of their PVAAS AGI being included in annual effectiveness rating.

Research Hypothesis 2: Teachers experience a change in the instructional practices they are expected to utilize as a result of their school's PVAAS AGI.

Research Hypothesis 3: Teachers experience staffing changes as a result of their school's PVAAS AGI.

Research Hypothesis 4: Teachers experience scheduling changes as a result of their school's PVAAS AGI.

Research Hypothesis 5: Teachers perceive a shift in the amount of support provided to students as a result of changes in the PVAAS AGI.

Research Hypothesis 6: Teachers do not believe that their PVAAS AGI score is solely dependent on teacher efficacy.

Research Hypothesis 7: Teachers have an incomplete understanding of how the PVAAS AGI score is calculated.

Data collection. Prior to collecting data in this phase, the proper steps were taken to obtain Internal Review Board (IRB) approval from Drexel University. A pilot survey was created and provided via email with an explanation of the purpose of the research to the Superintendent of the school district, for the purpose of obtaining approval to pilot the survey at elementary schools in the school district with the goal of eventually providing the finalized survey to PVAAS eligible teachers in grades six through eight in the school district. After receiving approval from the Superintendent to pilot and conduct the survey, the researcher communicated via email with the two principals of the middle schools in which the survey was intended to be conducted to obtain their approval for requesting that their teachers participate in the research survey. After receiving initial approval from the middle school principals, the four elementary school principals were contacted via email to obtain approval for piloting the survey with their PVAAS eligible teachers in grades four through five. Three of the four principals provided approval to pilot the survey. After receiving this permission, the pilot survey was distributed to teachers in grades four through five via email using Google forms. This data was collected and it was determined that all of the survey questions are unambiguous and that usable data could be gathered from the results. The forms for IRB approval were then completed and submitted to the IRB committee. After receiving IRB approval, a formal letter describing the research (Appendix B) and the finalized survey (Appendix E) was sent to the Superintendent of the school district requesting permission to provide the survey to their PVAAS eligible teachers of mathematics and English language arts in grades six through eight. Upon receiving approval from the Superintendent, the email addresses of each PVAAS eligible teacher in grades six through eight was collected and the survey, with consent disclosure, was supplied via email using Google forms from the researcher's private Google mail account (Appendix E). The data was then secured in this password protected account.

Data analysis procedures. Analyzing the likert scale data involved determining the mean score for each of the questions in the survey. Additionally, the percentage of participants selecting each scale point for each question was calculated to determine if this information can contribute to a broader understanding of the survey results.

Timeline

Figure 3: Research Timeline.

Ethical Considerations

There are three areas to be mindful of ethical considerations when conducting a quantitative research design both with and without human subjects: data collection, data analysis and data reporting. In the data collection process it is important to ensure that a large enough sample size is collected. The sample size should be greater than or equal to thirty data points. The first two phases of the study will include data points from 260 schools in the state of Pennsylvania, satisfying the sample size constraints. While it was not possible to obtain thirty responses in the third phase, the explanations provided with the responses yielded greater insight into the data obtained. It was necessary to obtain permission to survey the mathematics and English language arts teachers from the Superintendent of the surveyed school district, to provide accurate documentation to the survey participates of the type of study they are participating in, obtain permission to include their data, and ensure that their data remains anonymous. When analyzing the data, it was important to ensure accuracy, particularly when extracting and transcribing data from the data tables provided by PVAAS. The researcher must also keep in mind that the outcomes of data analysis for the first two phases of the study does not necessary imply cause, only correlation. All data analysis in this study was reported, including that which does not support any correlation.

Summary

In Pennsylvania, fifteen percent of a teacher's effectiveness rating is determined by the PVAAS AGI calculated for the teacher's school. The accuracy of the PVAAS AGI score calculation has become a subject of national debate and crucial to teachers wanting to ensure their professional status is earned. This research sought to determine if there are any indicators that the PVAAS AGI may not be a valid measure of teacher efficacy. Such indicators include the level of consistency of PVAAS AGI scores in English language arts and mathematics, correlations between PVAAS AGI scores and demographics, and the changes perceived by teachers to occur as the result of the PVAAS AGI that are not directly related to a teacher's individual effectiveness.

Chapter 4: Findings and Results

The purpose of this research was to determine if the Pennsylvania Value-Added Assessment System Average Growth Index (PVAAS AGI) scores for teachers and schools provide a valid and consistent assessment of teacher effectiveness and if teachers' perceptions indicate that the validity and consistency of the PVAAS AGI score included in the Classroom Rating Tool is important. Data collection and analysis were guided by three research questions:

1. To what extent do PVAAS AGI scores assigned to a school change statistically over time?
2. How does the PVAAS AGI scores assigned to a school relate to the percentage of economically disadvantaged, learning disabled, English language learners, and minorities in the school?
3. To what extent do teacher perceptions of the impact of PVAAS AGI indicate that the validity of the PVAAS AGI calculated for their school is important?

A mixed ANOVA was utilized to determine how PVAAS AGI scores change statistically over time in mathematics and English language arts. Two hundred and sixty middle schools serving grades six through eight were first categorized by achievement level in mathematics, based on data from the base year, 2013. Schools with an average Normal Curve Equivalent (NCE) score of 60% or above were categorized as Level 1 schools. Schools with an average NCE scores greater than or equal 50% and less than 60% were categorized as Level 2. Level 3 schools had average NCE scores that were greater than or equal to 40% and less than 50%. Schools were categorized as Level 4 schools if they had an average NCE scores greater than or equal to 30% and less than 50%. Finally, Level 5 schools had average NCE scores that were greater than or equal to 20% and less than 30%. The same process was utilized to categorize schools by achievement level in English language arts. None of the 260 schools had an average NCE score greater than 70% or less than 20% in either mathematics or English language arts.

In the mixed ANOVA design, achievement levels were treated as the between factors and the PVAAS AGI scores for 2013-2016 were treated as the within factors. The goal of this test was to determine if there was any difference in mean gain scores between schools in different achievement levels and four years of annual PVAAS AGI scores. This test was performed twice, once using achievement level data and PVAAS AGI in mathematics and once using achievement level data and PVAAS AGI in English language arts.

A Pearson correlation test was used to determine if any correlations existed between PVAAS AGI in mathematics and demographic variables of each school as well as between PVAAS AGI in English language arts and the demographic variables of each school. Scatterplot graphs were generated for each test to represent the distribution of PVAAS AGI scores in mathematics and English language arts as they relate to the percentage of students in each demographic category. The four demographic categories tested included economically (ED) disadvantaged, special education students with individualized education plans (IEP), limited English proficient (LEP), and minority (MIN). Students classified as limited English proficient are also sometimes referred to in this analysis as English language learners.

A likert scale survey was utilized to determine teacher perceptions of change in the school environment resulting from PVAAS AGI scores being included in their Classroom Rating Tool (annual teacher evaluation), and whether teachers perceive PVAAS AGI calculations to be valid. The survey was distributed to two middle schools, within the same school district, serving grades six through eight. PVAAS eligible teachers were invited to complete the survey electronically via Google forms.

Results and Interpretations

A mixed ANOVA design was utilized to determine the extent to which PVAAS AGI scores in mathematics and English languages arts change over time. Pearson correlation tests were used to identify correlations between PVAAS AGI and demographic variables in school buildings. A survey utilizing the likert scale was administered to teachers in two middle school buildings to determine the importance of consistency and validity of the PVAAS AGI scores included in the teachers' Classroom Rating Tool.

PVAAS AGI Statistical Changes over Time

During the data collection process it was discovered that there were a number of discrepancies in the reporting of grade levels for each school. The first phase of data collection involved using the search tool (PVAAS, 2016) in September of 2016, when the 2015 data was still available in the search tool (PVAAS only publishes the previous year's data in the search tool), to select all of the schools that served grades six through eight. The scatterplot tool was then utilized to collect information for this set of schools from 2013 and 2014. Four schools were not available in the scatterplot tool. Unable to secure four years of data for these schools, they were not included in the final analysis. In October of 2016, after the 2016 data was entered into the search tool, a similar search for schools serving grades six through eight was performed. This search yielded a set of schools that either excluded schools from the previous year or included schools that had not been included the previous year. Upon further investigation, it was found that multiple schools had their grade levels listed incorrectly in the PVAAS search tool. Table 6 represents the schools that were misclassified in the 2016 search tool. It is unclear whether this discrepancy only effects the grade level identification of the school on the website or if the misclassification of the schools also resulted in miscalculation due to including the incorrect set of testing data. Several schools were also listed with incorrect names or not reported at all. Table 6 represents the number of schools serving grades six through eight that were identified as serving incorrect grade levels. Only schools for which it was possible to obtain four years of data from the PVAAS public website were considered in the analysis. The grade levels for schools that were not listed as serving grades six through eight on the PVAAS website in 2016 were independently verified using the individual school district websites before their data was included in the analysis. A total of 29 schools or 11% of the schools included in the analysis were incorrectly identified by grade level in 2016.

Table 6

PVAAS grade level reporting inaccuracies in 2016

Number of Schools Serving Grades 6 to 8	Incorrect Grade Levels Listed
2	4-8
16	5-8
8	7-8
2	6-11
1	6-12

A Mixed ANOVA design was utilized to determine how the PVAAS AGI scores in mathematics and English language arts of Pennsylvania middle schools hosting grades six through eight changed over time. Schools were categorized into five levels (between factors) of student performance using the school's average normal curve equivalent scores and tested using four years of PVAAS AGI data (within factors) in mathematics and English language arts from the years 2013 to 2016. The first product of this analysis utilizing PVAAS AGI data for mathematics and performed using the Statistical Package for Social Sciences (SPSS) revealed means and standard deviations in PVAAS AGI scores in mathematics for schools in each level. Table 7 demonstrates that PVAAS AGI scores for mathematics in grades six through eight have, on average, decreased consistently from 2013 to 2016 across all five performance levels.

Table 7

Univariate statistics for mathematics variables in analysis (standard errors)

		Means			
Group	N	2013	2014	2015	2016
Level 1	36	2.87 (4.05)	1.4 (4.42)	1.63 (5.67)	-1.83 (10.98)
Level 2	147	1.85 (4.23)	-.8 (4.72)	-.67 (4.2)	.11 (4.4)
Level 3	60	1.39 (3.59)	-.21 (3.86)	-0.57 (4.78)	.04 (3.76)
Level 4	14	2.9 (2.75)	.97 (5.15)	-2.84 (4.96)	-3.07 (4.4)
Level 5	3	-.91 (1.79)	-2.34 (1.79)	-1.41 (1.17)	-1.64 (1.7)

The second part of the analysis provided results of the mixed ANOVA. Mauchly's Test was highly significant, W = 0.858, χ^2 (5) =38.75, p < .000, violating the null hypothesis that the matrix does not have equal variances and equal covariances. Therefore, the Greenhouse- Geisser and Hunh-Feldt epsilon corrections can be used to ensure that the F-test does not result in an inflation of Type 1 Errors. The corresponding corrective coefficients were: Greenhouse-Geisser $\varepsilon = .911$ and Hunyh-Feldt $\varepsilon = .937$. However, given that even with each correction the F value remains the same, it was not necessary to use either of the corrections. This analysis revealed that time had a negative effect on PVAAS AGI scores in mathematics. The results are presented in Table 8.

Table 8

Repeated measures analysis of variance for mathematics within-subjects effects

Effect	*MS*	*df*	*F*	*P*	η_p^2
Time	87.61	3	5.21	< .001	.02
Time x Level	52.32	12	2.84	< .001	.043
Error	16.8	765			

Analysis of the between-subject effects revealed that there was not a statistically significant difference in variance between groups and within groups. Therefore, the performance level of a school does not seem influence the PVAAS AGI score in mathematics over time (see table 9).

Table 9

Repeated measures analysis of variance for mathematics between-subjects effects

Source	*MS*	*df*	*F*	*P*	η_p^2
Level	41.436	4	1.069	< .372	.016

Error	16.8	765

The second mixed ANOVA analysis utilized PVAAS AGI data in English language arts for schools categorized by performance level. The first output of this analysis revealed means and standard deviations in PVAAS AGI scores in English language arts for schools in each performance level. Table 10 demonstrates that PVAAS AGI scores for mathematics in grades six through eight have, on average, decreased consistently from 2013 to 2016 across all five performance levels.

Table 10

Univariate statistics for English language arts variables in analysis (standard errors)

		Means			
Group	N	2013	2014	2015	2016
Level 1	29	1.24 (3.5)	1.27 (3.34)	2.39 (3.84)	-2.13 (2.82)
Level 2	141	1.11 (2.77)	-3.2 (3.02)	.11 (3.82)	-0.9 (3.35)
Level 3	73	1.81 (2.5)	-.49 (2.86)	-1.05 (3.96)	0.25 (3.41)
Level 4	16	2.51 (1.55)	-.95 (1.6)	-1.92 (2.9)	-1.74 (2.6)
Level 5	1	0.88	-1.03	-5.9	-1.86

The second part of the analysis provided results of the mixed ANOVA. Mauchly's Test was highly significant, W = 0.908, χ^2 (5) =24.476, p < .000 violating the null hypothesis that the matrix does not have equal variances and equal covariances. Therefore, the Greenhouse- Geisser and Hunh-Feldt epsilon corrections can be used to ensure that the F-test does not result in an inflation of Type 1 Errors. The corresponding corrective coefficients were: Greenhouse-Geisser ε = .944 and Hunyh-Feldt ε = .971 . However, given that even with each correction the F value remains the same, it was not necessary to use either of the corrections. This analysis revealed that time had a negative effect on PVAAS AGI scores in mathematics (see table 11).

Table 11

Repeated measures analysis of variance for English language arts within-subjects effects

Effect	*MS*	*df*	*F*	*P*	η_p^2
Time	36.42	3	4.63	.003	.02
Time x Level	45.06	12	5.73	.000	.08
Error	7.86	722.35			

Analysis of the between-subject effects for English language arts revealed that there was not a statistically significant difference in variance between groups and within groups. Therefore, the performance level of a school did not seem to influence the PVAAS AGI score in English language arts over time (see table 12).

Table 12

Repeated measures analysis of variance for English language arts between-subjects effects

Source	*MS*	*df*	*F*	*P*	η_p^2
Level	20.41	4	.44	.334	.02
Error	17.75	255			

The finding of greatest interest from this analysis was that the means for schools across all performance levels and subject areas significantly declined over the four year period. It is important to note when interpreting these findings that in 2015 the Pennsylvania State Department of Education published a new test aligned with the new Pennsylvania Common Core Standards. Figures 4 and 5 provide a graphical representation of how the mean gain scores changed over time. It is apparent from Figure 2 that the PVAAS AGI mathematics and English language arts scores fell significantly overall from 2013 to 2016. However, the mean gains of levels 1, 2, and 5 in both mathematics and English language arts experienced an increase in 2015.

Figure 4. Mean gain score changes over time in PVAAS AGI for mathematics.

Figure 5. Mean gains over time in PVAAS AGI for English language arts.

Relationships between PVAAS AGI Scores and Demographics

PVAAS AGI scores are calculated using the Normal Curve Equivalent (NCE) of each student. Therefore, the first set of Pearson correlation tests conducted were between the average NCE for each school and the percentage of economically disadvantaged, special education, limited English proficient, and minority students in the school. This simple calculation was calculated using Microsoft Excel. Then, SPSS was utilized to calculate an individual Pearson correlation test for the mathematics and English language arts PVAAS AGI scores for each school in 2015 and 2016 and the percentage of students within each school composing each of the demographic categories: economically disadvantaged, learning disabled, English language learner, and minority. The Pearson correlation r values were calculated and a scatterplot of the data and best fit line was generated.

2015 PVAAS AGI in Mathematics

The first set of Pearson correlation tests for mathematics in 2015 analyzed the relationship between the average NCE score for each school in mathematics and the percentage of economically disadvantaged, special education, limited English proficiency, and minority students in 2015. Table 13 represents the relationship between the average NCE score for each school and each of the demographic variables.

Table 13

Pearson correlations between average Normal Curve Equivalent scores in mathematics and demographic variables in 2015

Demographic Variable	Pearson's r
Economically Disadvantaged	-0.86
Special Education	-0.49
Limited English Proficiency	-0.45
Minority	-0.6

There was a strong negative correlation between average NCE scores for each school and the percentage of economically disadvantaged students. Weak negative correlations were identified between average NCE scores and the percentage of special education and limited English proficiency students. A moderate negative correlation was identified to exist between the average NCE score for each school and the percentage of minority students.

The second set of Pearson tests for mathematics in 2015 analyzed the relationship between PVAAS AGI in mathematics for each school and the percentage of economically disadvantaged, special education, limited English proficient, and minority students in each school. There was a very weak negative correlation between the PVAAS AGI mathematics scores for each school and the percentage of economically disadvantaged students in the school, $r = -.201$, $n=260$, $p=.001$. Figure 6 shows the distribution of the PVAAS AGI mathematics scores per school as related to percent economically disadvantaged in each school and line of best fit.

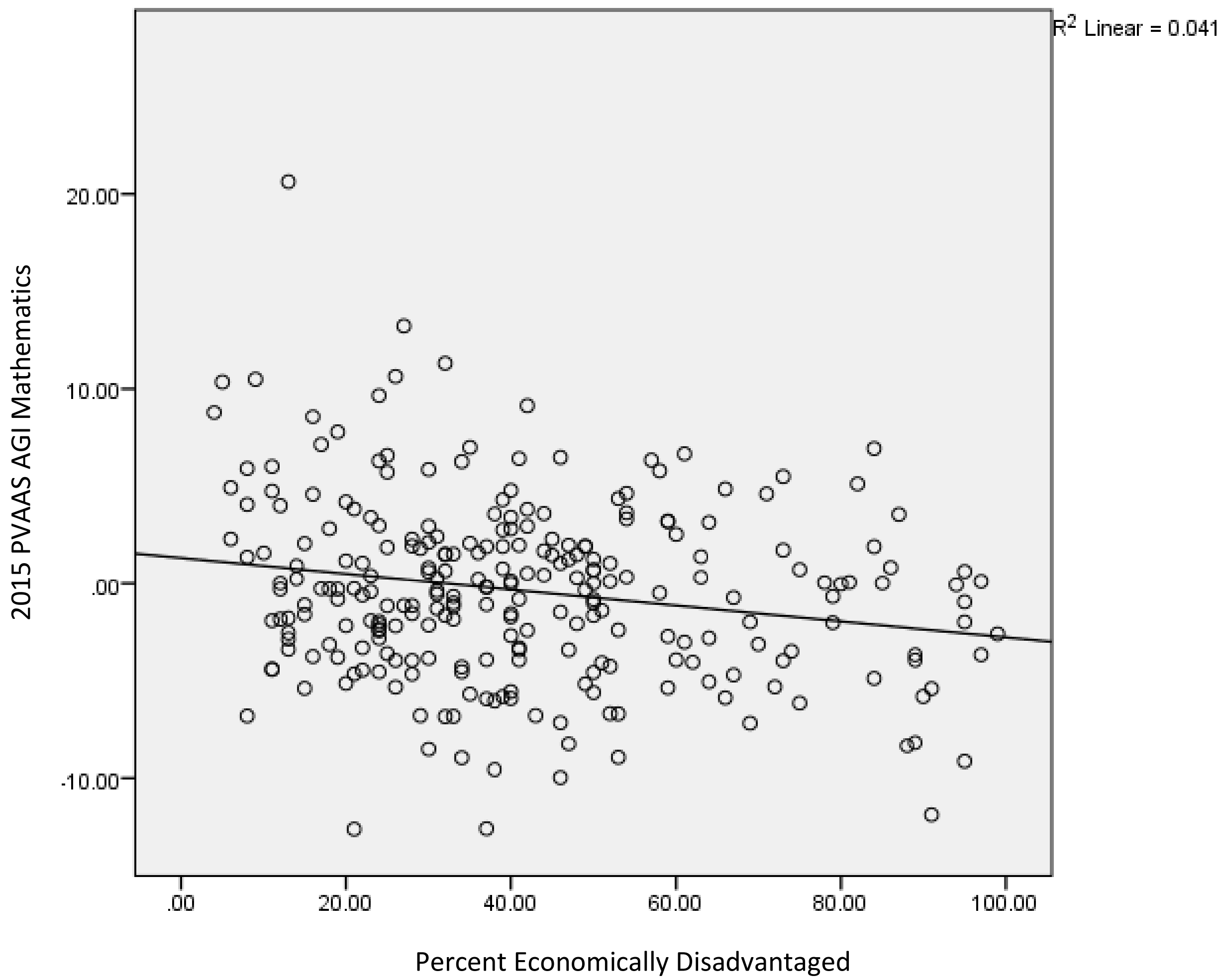

Figure 6. PVAAS AGI for mathematics related to percent economically disadvantaged.

There was a very weak negative correlation between the PVAAS AGI mathematics score for each school building and the percentage of students requiring special education services in each building, r = -.166, n= 260, p = .004. Figure7 represents the distribution of PVAAS AGI mathematics scores and the line of best fit.

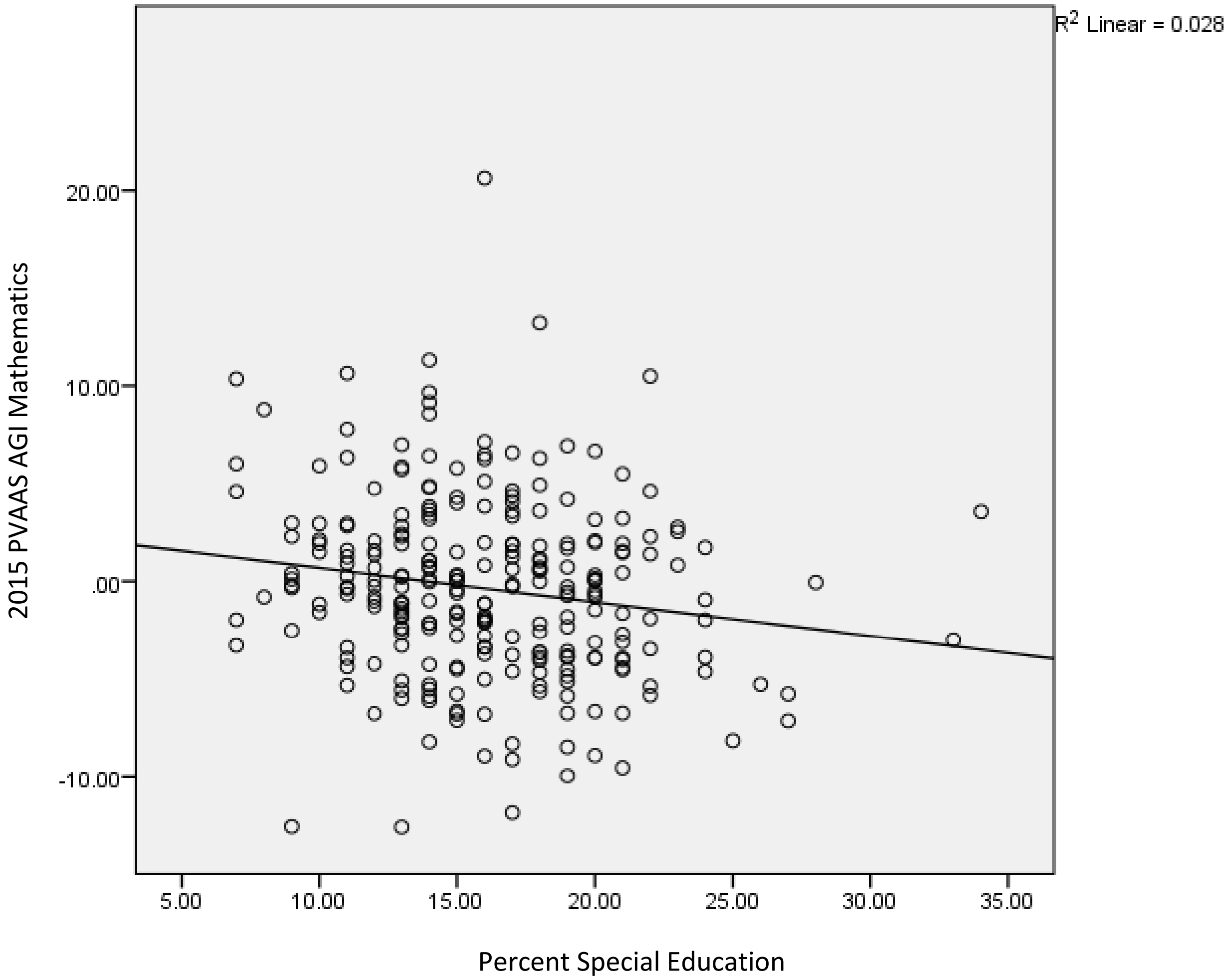

Figure 7. PVAAS AGI for mathematics related to percent of student needing special education services.

There was a very weak negative correlation between the PVAAS AGI mathematics scores for each school building and the percentage of students speaking English as a second language in each building, r = -.136, n= 260, p = .014. Figure 8 represents the distribution of PVAAS AGI mathematics scores and line of best fit.

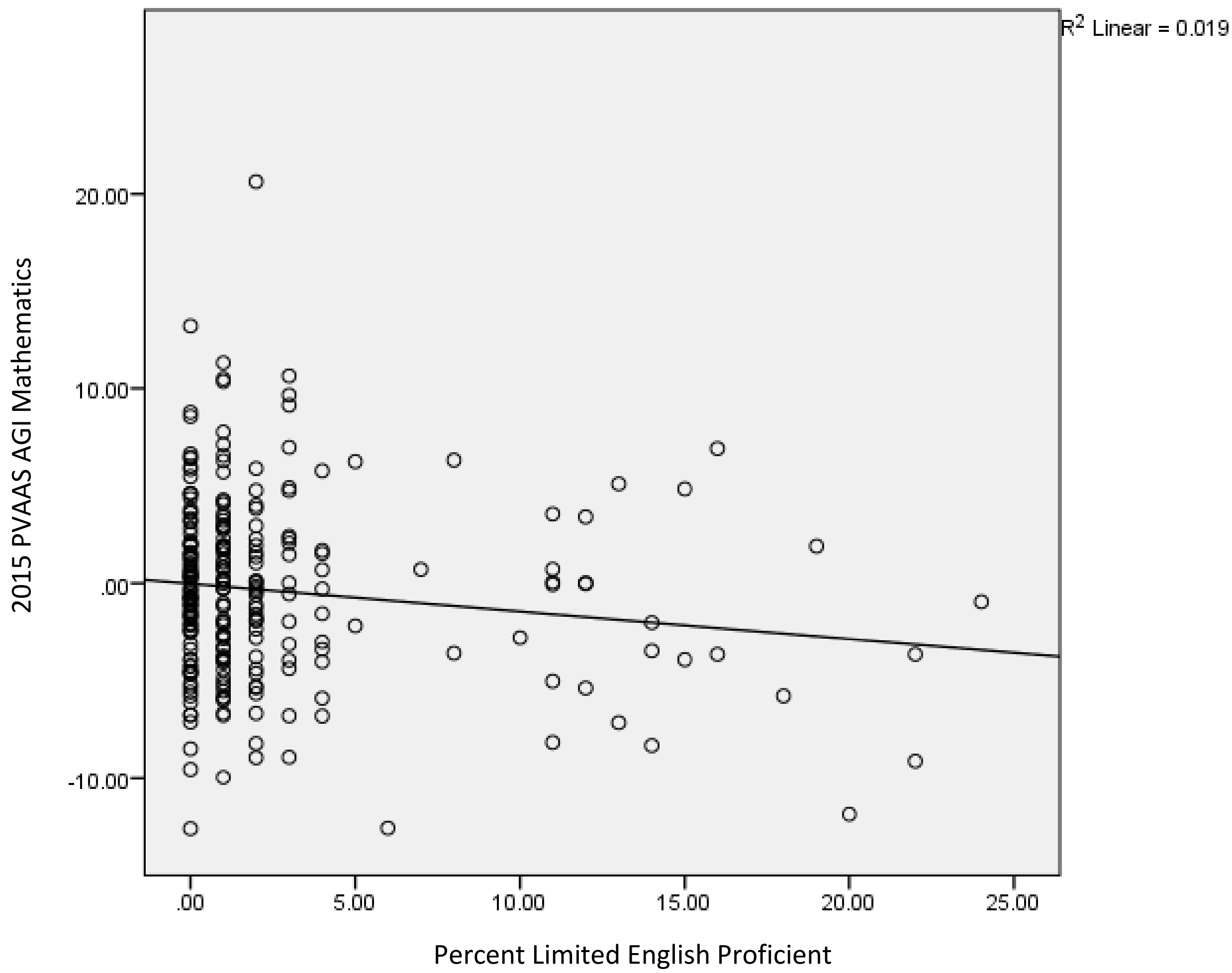

Figure 8. PVAAS AGI for mathematics related to percent limited English proficiency.

There was a very weak negative correlation between the PVAAS AGI mathematics scores for each school building and the percentage of student classified as minorities in each building, r = -.176, n =260, p = .002. Figure 9 represents the distribution of PVAAS AGI mathematics scores and line of best fit.

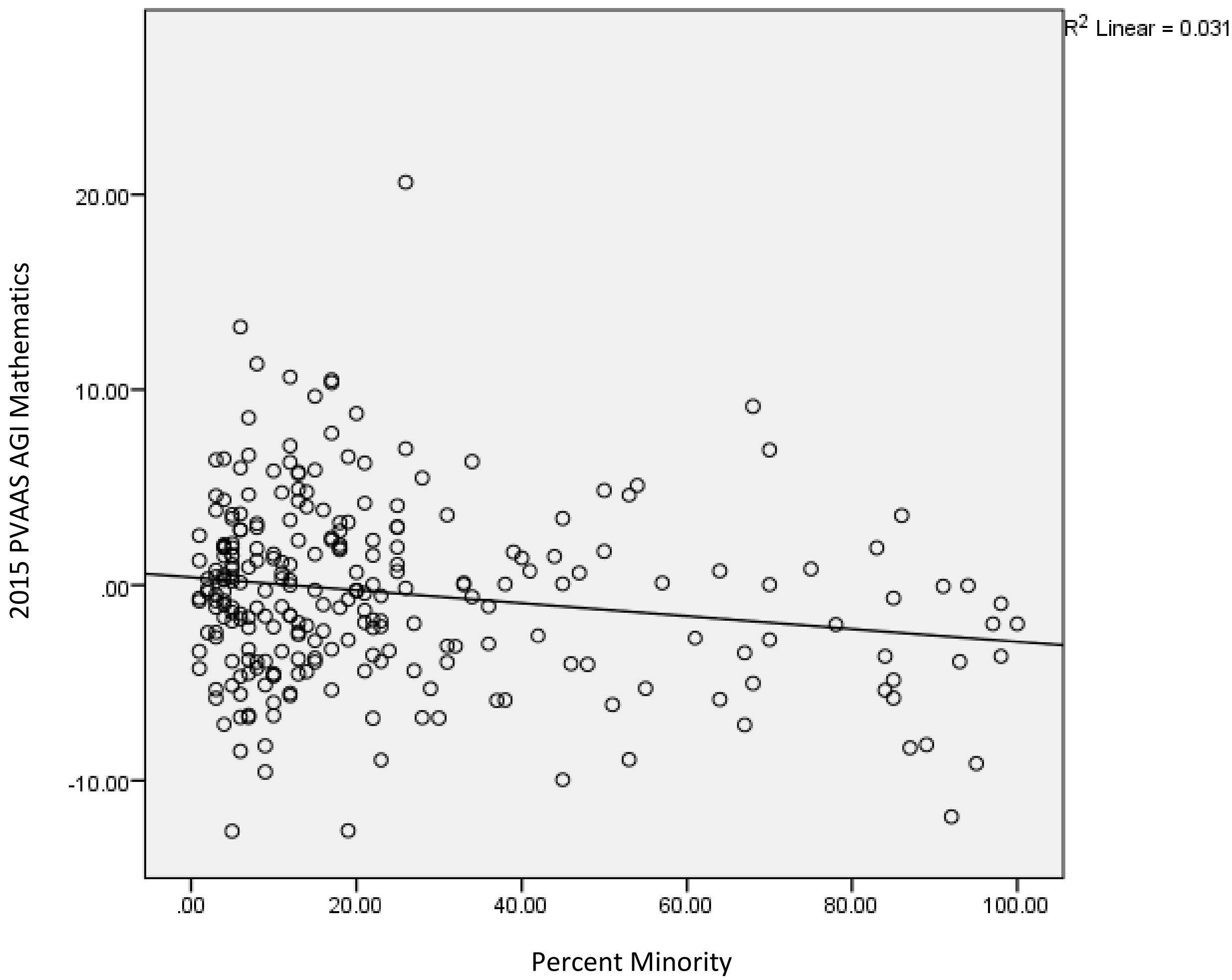

Figure 9. PVAAS AGI for mathematics related to percent minority.

2015 PVAAS AGI in English Language Arts

The first set of Pearson correlation tests identified relationships between the average NCE score of each school in English language arts and the percentage of economically disadvantaged, special education, limited English proficiency, and minority students in each school in 2015. Table 14 represents the relationship between the average NCE for each school and each of the demographic variables.

Table 14

Pearson correlations between average Normal Curve Equivalent scores in English language arts and Demographic Variables in 2015

Demographic Variable	Pearson's r
Economically Disadvantaged	-0.83
Special Education	-0.48
Limited English Proficiency	-0.42
Minority	-0.54

There was a strong negative correlation between average NCE scores for each school and the percentage of economically disadvantaged students. Weak negative correlations were identified between average NCE scores and the percentage of special education and limited English proficiency students. A moderate negative correlation was identified to exist between the average NCE score for each school and the percentage of minority students.

The second set of Pearson tests for English language arts in 2015 analyzed the relationship between PVAAS AGI in English language arts for each school and the percentage economically disadvantaged, special education, limited English proficient, and minority students in each school. There was a very weak positive correlation between the PVAAS AGI English language arts score for each school building and the percentage of students classified as economically disadvantaged in each building, $r = .28$, $n= 260$, $p = .000$. Figure 10 represents the distribution of PVAAS AGI English language arts scores and the line of best fit.

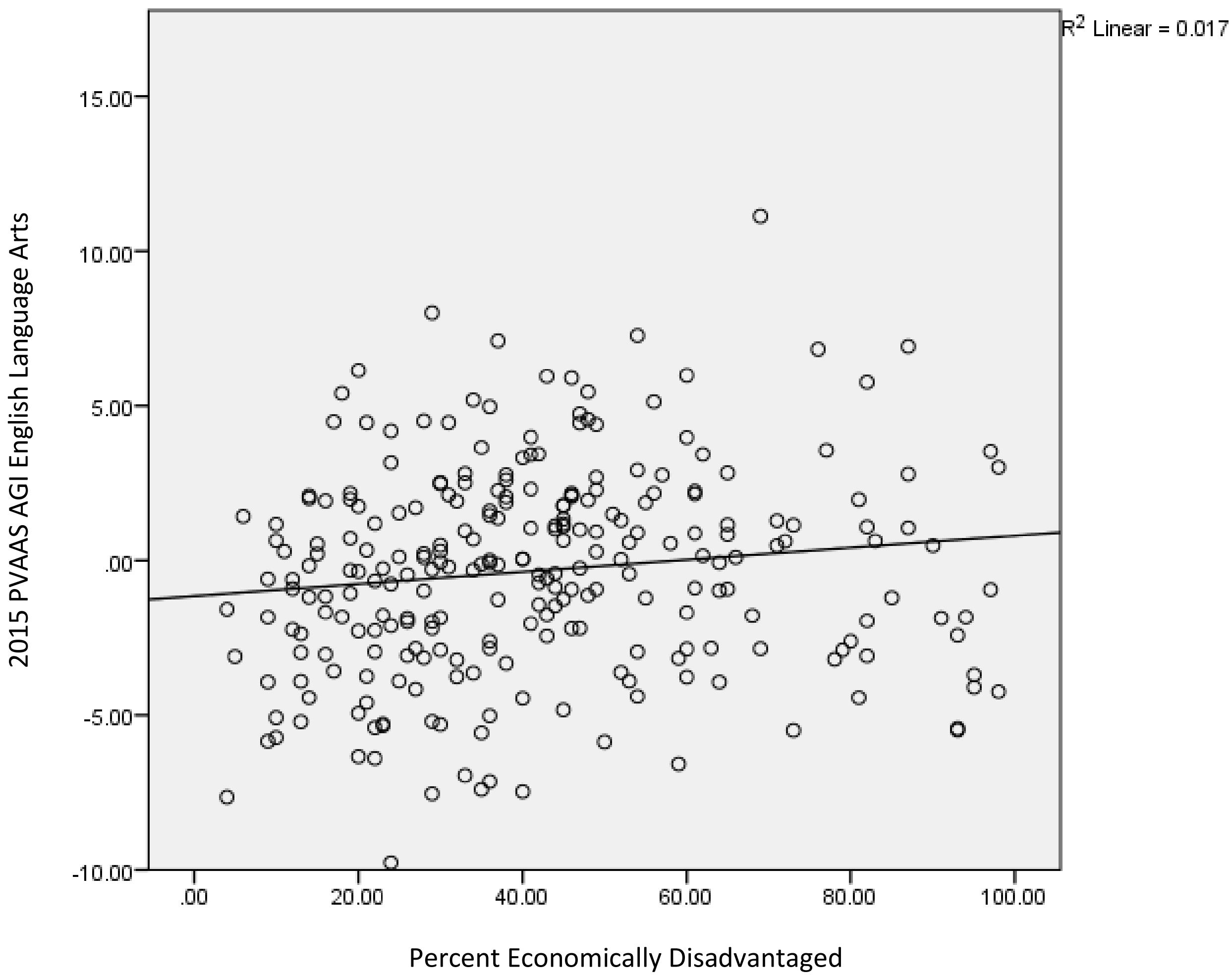

Figure 10. PVAAS AGI for English language arts related to percent economically disadvantaged.

There was a very weak negative correlation between the PVAAS AGI English language arts score for each school building and the percentage of students requiring special education services in each building, r = -.19, n= 260, p = .001. Figure 11 represents the distribution of PVAAS AGI English language arts scores and the line of best fit.

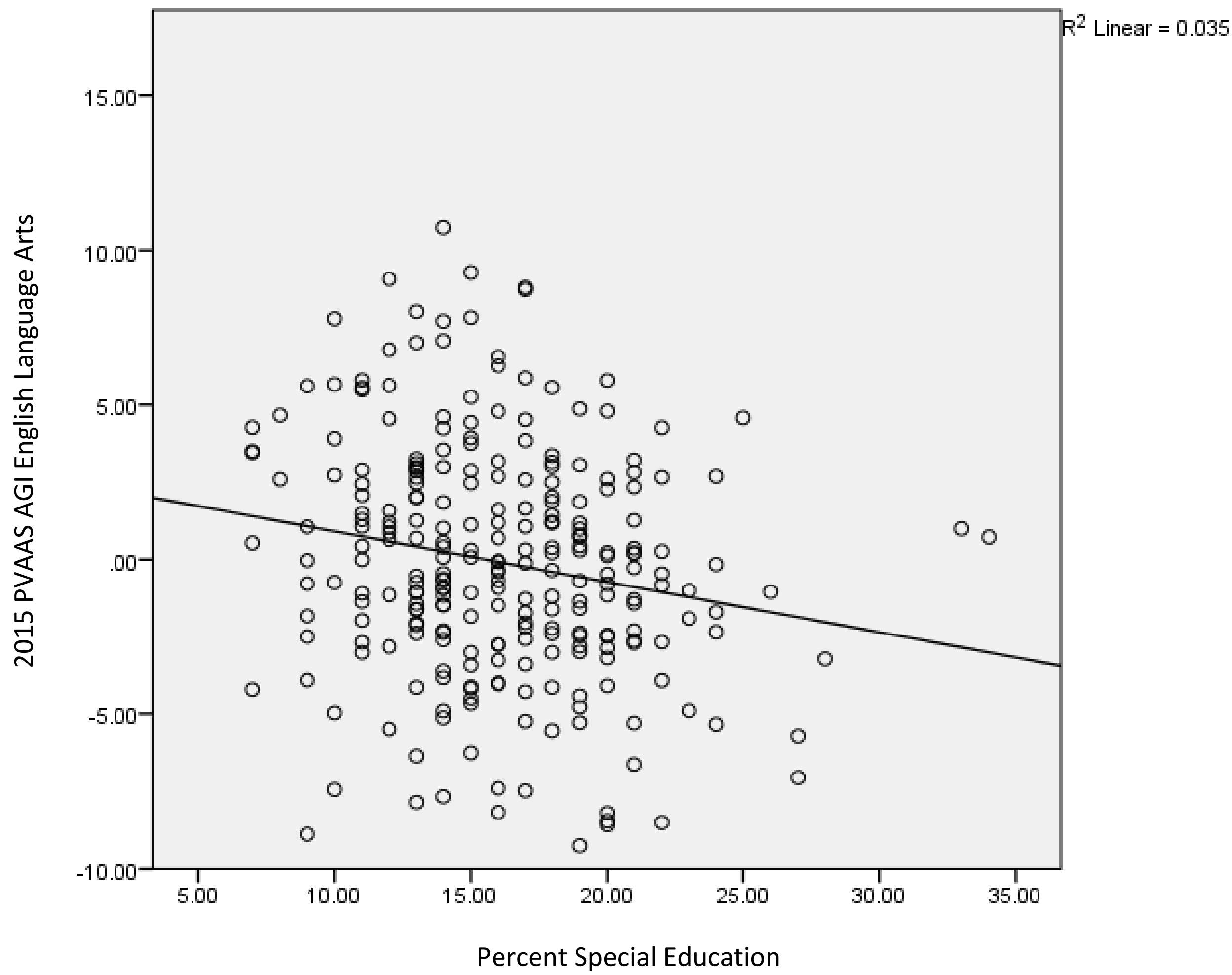

Figure 11. PVAAS AGI for English language arts related to percent needing special education services.

There was no significant correlation between the PVAAS AGI English language arts score for each school building and the percentage of students classified as limited English proficiency in each building, r = -.081, n= 260, p = .097. Figure 12 represents the distribution of PVAAS AGI English language arts scores and the line of best fit. However, in this case the best fit line is not indicative of a relationship due to the lack of significance in the correlation.

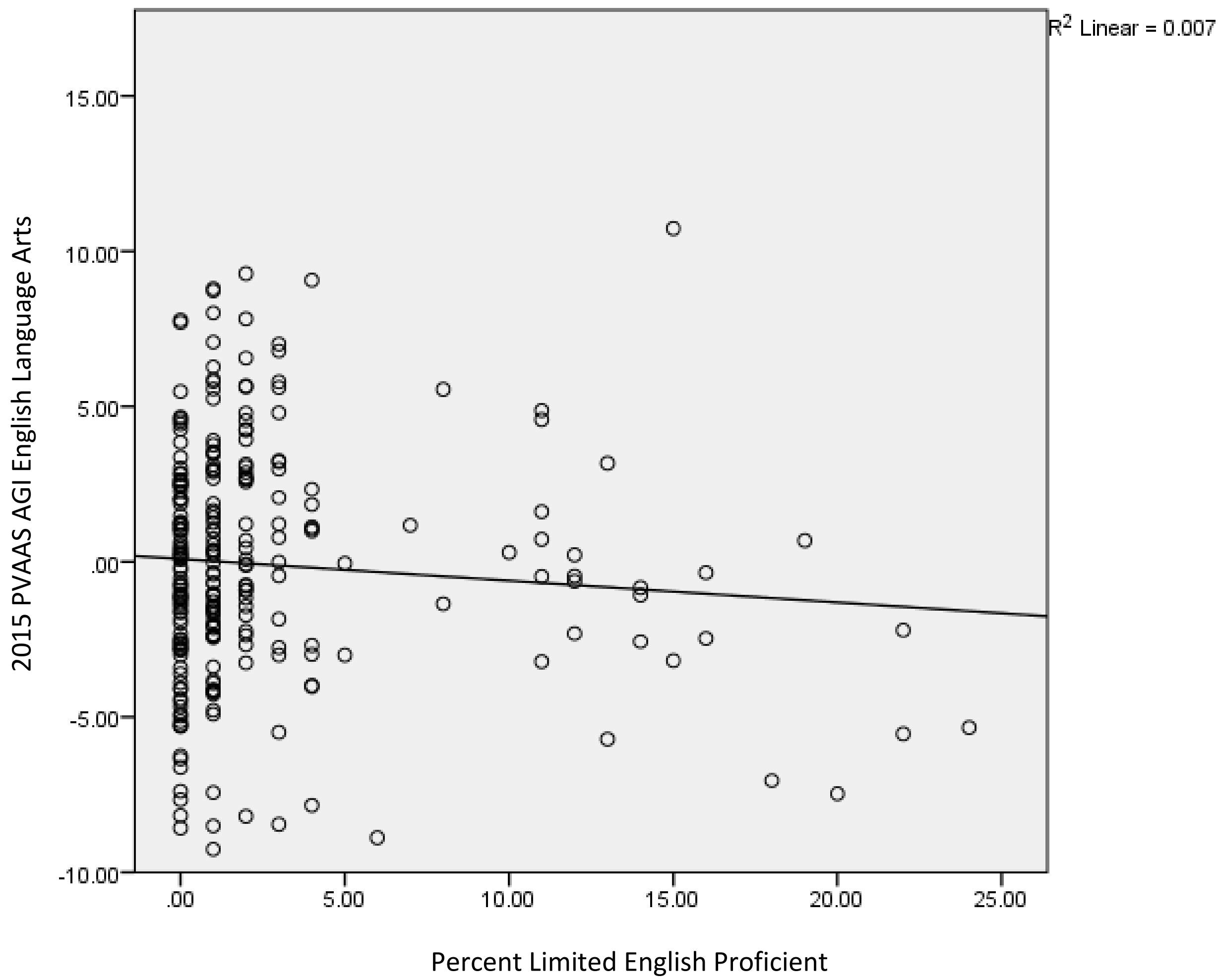

Figure 12. PVAAS AGI for English language arts related to percent limited English proficient.

There was a very weak negative correlation between the PVAAS AGI English language arts scores for each school building and the percentage of student classified as minorities in each building, r = -.115, n =260, p = .032. Figure 13 represents the distribution of PVAAS AGI English language arts scores and line of best fit.

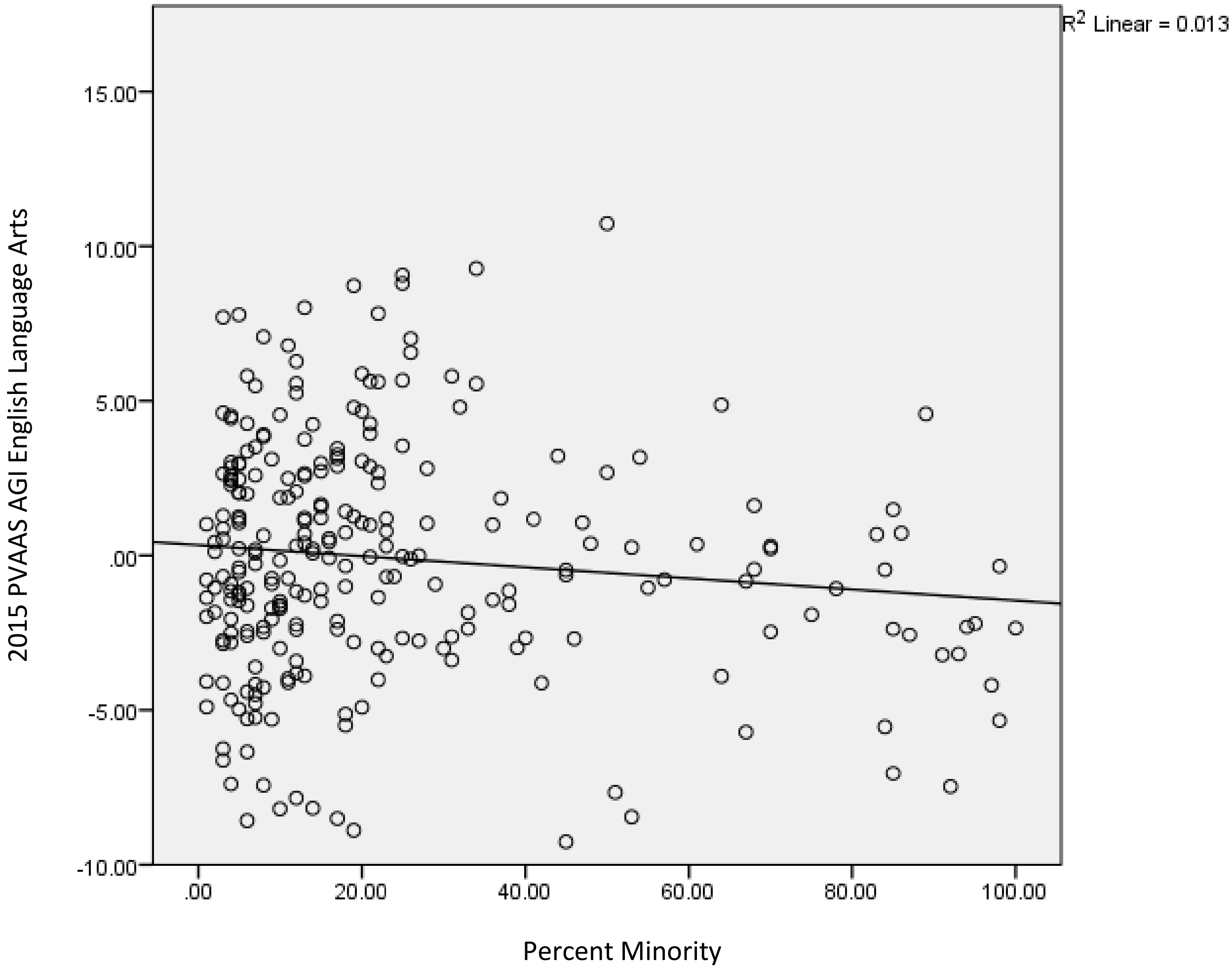

Figure 13. 2015 PVAAS AGI for English language arts related to percent minority.

2016 PVAAS AGI in Mathematics

The first set of Pearson tests for mathematics in 2016 analyzed the relationship between the average NCE with each school in mathematics and the percentage of economically disadvantaged, special education, limited English proficiency, and minority students in 2016. Table 15 represents that relationship between the average NCE in mathematics for each school and each of the demographic variables in 2016.

Table 15

Pearson correlations between average Normal Curve Equivalent scores in Mathematics and Demographic Variables in 2016

Demographic Variable	Pearson's r
Economically Disadvantaged	-0.85
Special Education	-0.51
Limited English Proficiency	-0.43
Minority	-0.61

There was a strong negative correlation between average NCE scores for each school and the percentage of economically disadvantaged students. Moderate negative correlations were identified between average NCE scores and the percentage of special education and minority students. A weak negative correlation was identified to exist between the average NCE score for each school and the percentage of limited English proficiency students.

The second set of Pearson tests for mathematics in 2016 analyzed the relationship between PVAAS AGI in mathematics for each school and the percentage of economically disadvantaged, special education, limited English proficient, and minority students in each school. There was no significant correlation between the 2016 PVAAS AGI mathematics scores for each school and the percentage of economically disadvantaged students in the school, r = -.058, n=260, p=.174. Figure 14 shows the distribution of the 2016 PVAAS AGI mathematics scores and line of best fit. However, in this case the best fit line is not indicative of a relationship due to the lack of significance in the correlation.

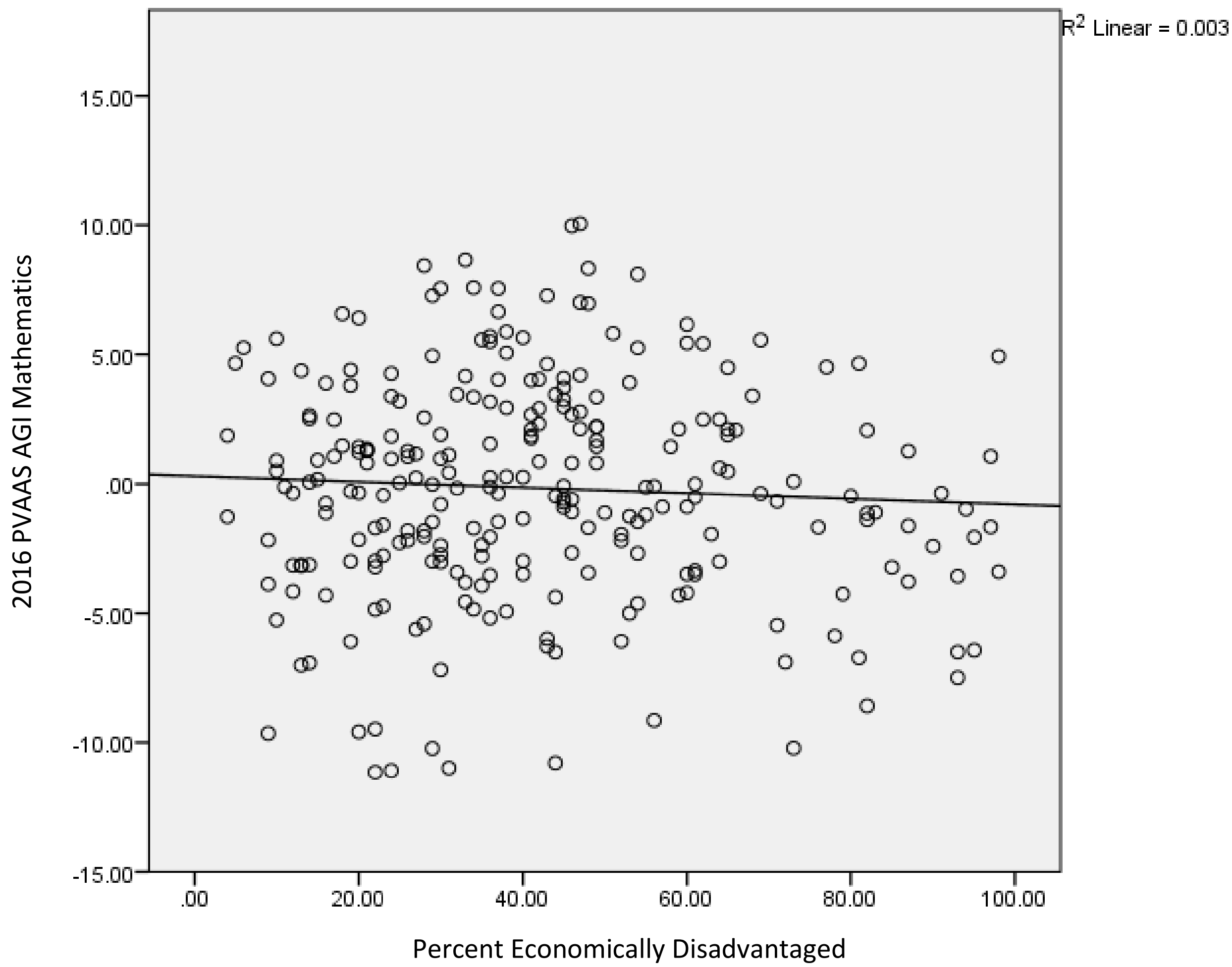

Figure 14. 2016 PVAAS AGI for mathematics related to percent economically disadvantaged.

There was no significant correlation between the 2016 PVAAS AGI mathematics score for each school building and the percentage of students requiring special education services in each building, r = -.077, n= 260, p = .109. Figure 15 represents the distribution of 2016 PVAAS AGI mathematics scores and the line of best fit. However, in this case the best fit line is not indicative of a relationship due to the lack of significance in the correlation.

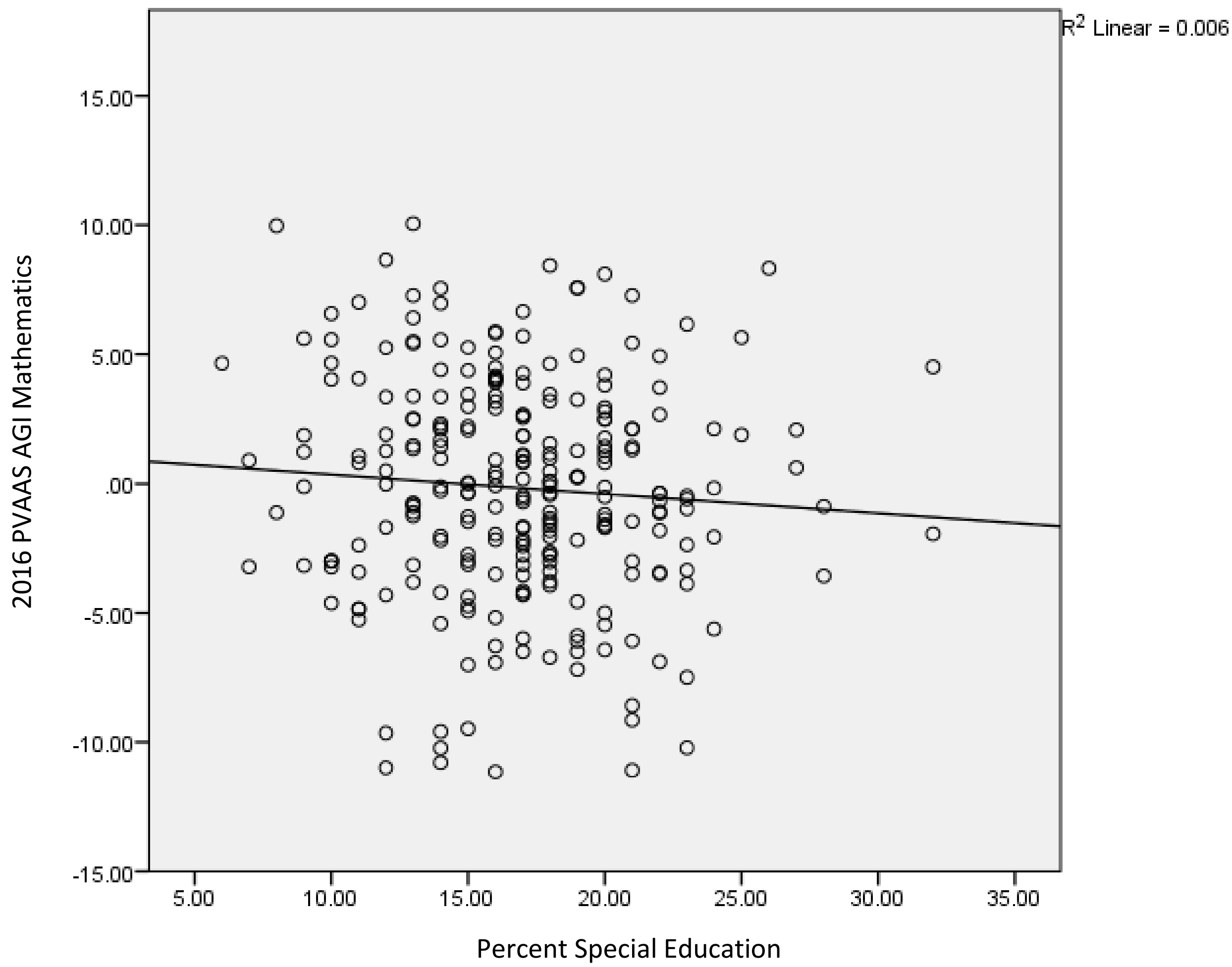

Figure 15. 2016 PVAAS AGI for mathematics related to percent needing special education services.

There was a very weak correlation between the 2016 PVAAS AGI mathematics score for each school building and the percentage of students classified as limited English proficiency in each building, r = -.129, n= 260, p = .019. Figure 16 represents the distribution of 2016 PVAAS AGI mathematics scores and the line of best fit.

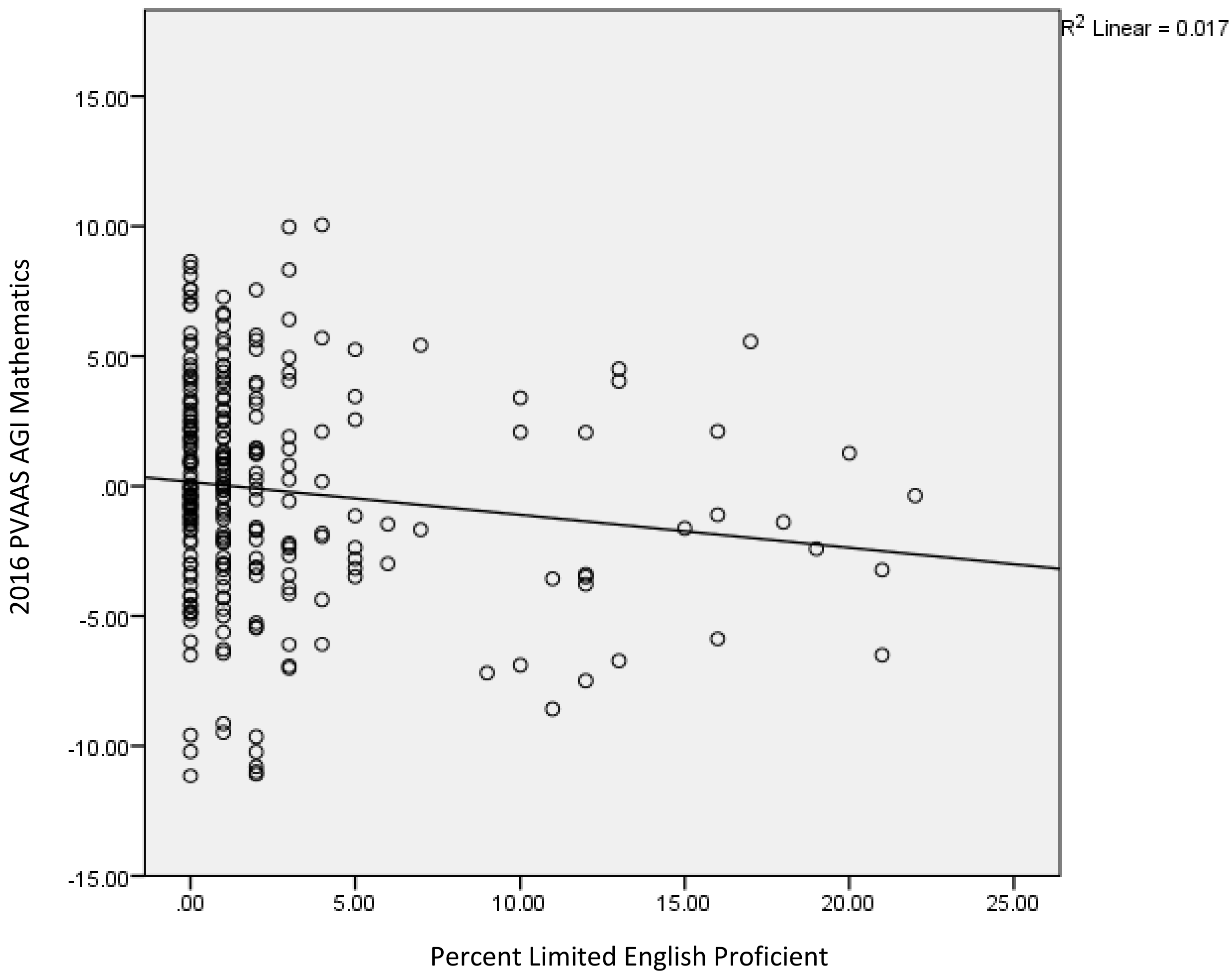

Figure 16. 2016 PVAAS AGI for mathematics related to percent limited English proficient.

There was a very weak negative correlation between the PVAAS AGI mathematics scores for each school building and the percentage of student classified as minorities in each building, r = -.182, n =260, p = .002. Figure 17 represents the distribution of PVAAS AGI mathematics scores and line of best fit.

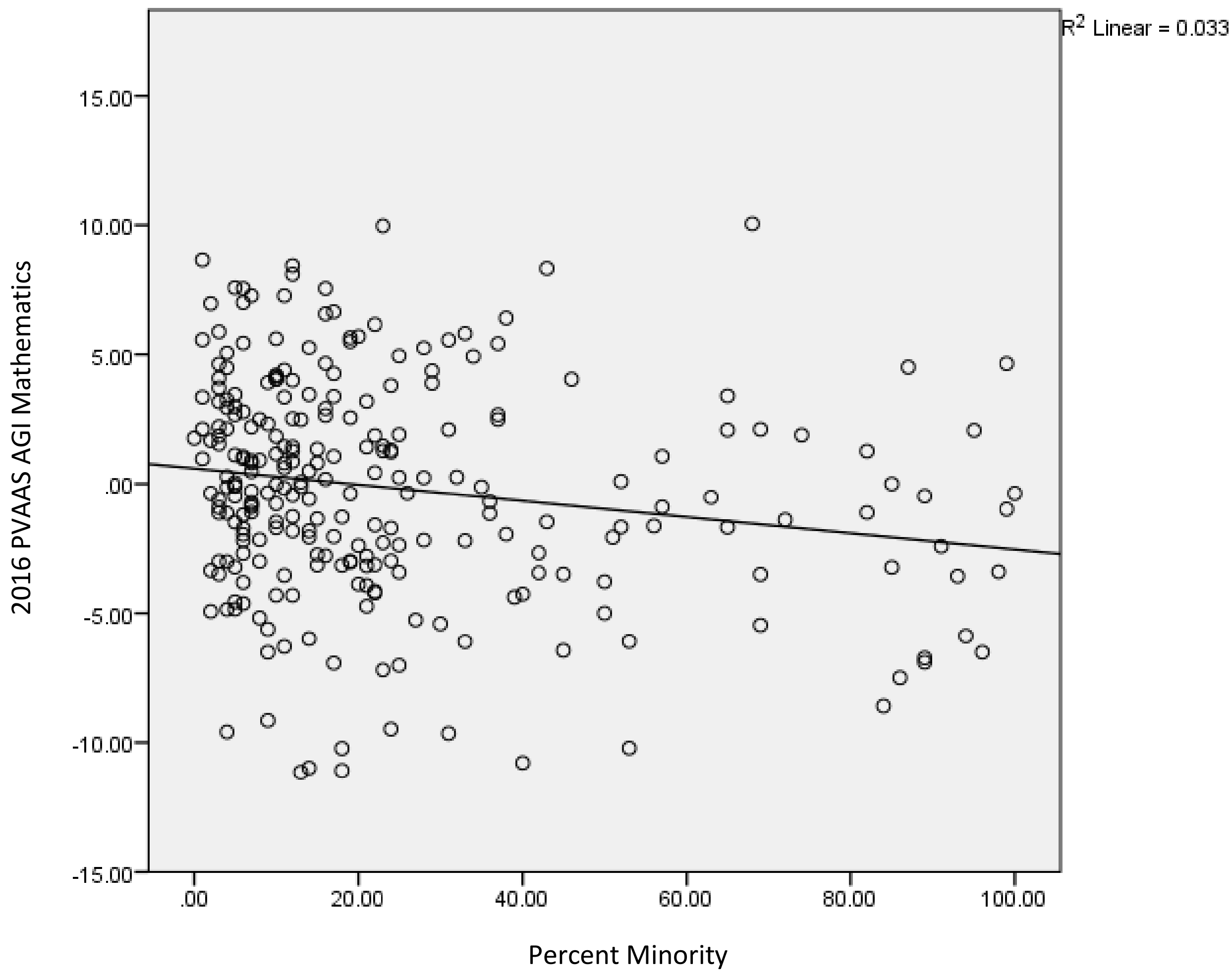

Figure 17. 2016 PVAAS AGI for mathematics related to percent minority.

2016 PVAAS AGI in English Language Arts

The first set of Pearson tests for English language arts in 2016 analyzed the relationship between the average NCE for each school in English language arts and the percentage of economically disadvantaged, special education, limited English proficiency, and minority students in 2016.

Table 16

Pearson correlations between average Normal Curve Equivalent scores in English language arts and demographic variables in 2016

Demographic Variable	Pearson's r
Economically Disadvantaged	-0.84
Special Education	-0.5
Limited English Proficiency	-0.44
Minority	-0.6

There was a strong negative correlation between average NCE scores for each school and the percentage of economically disadvantaged students. Moderate negative correlations were identified between average NCE scores and the percentage of special education and minority students. A weak negative correlation was identified to exist between the average NCE score for each school and the percentage of limited English proficiency students.

The second set of Pearson tests for English language arts in 2016 analyzed the relationship between PVAAS AGI in mathematics for each school and the percentage economically disadvantaged, special education, limited English proficient, and minority students in each school. There was a very weak positive correlation between the 2016 PVAAS AGI English language arts scores for each school and the percentage of economically disadvantaged students in the school, $r = -.13$, $n=260$, $p=.018$. Figure 18 shows the distribution of the 2016 PVAAS AGI English language arts scores and line of best fit.

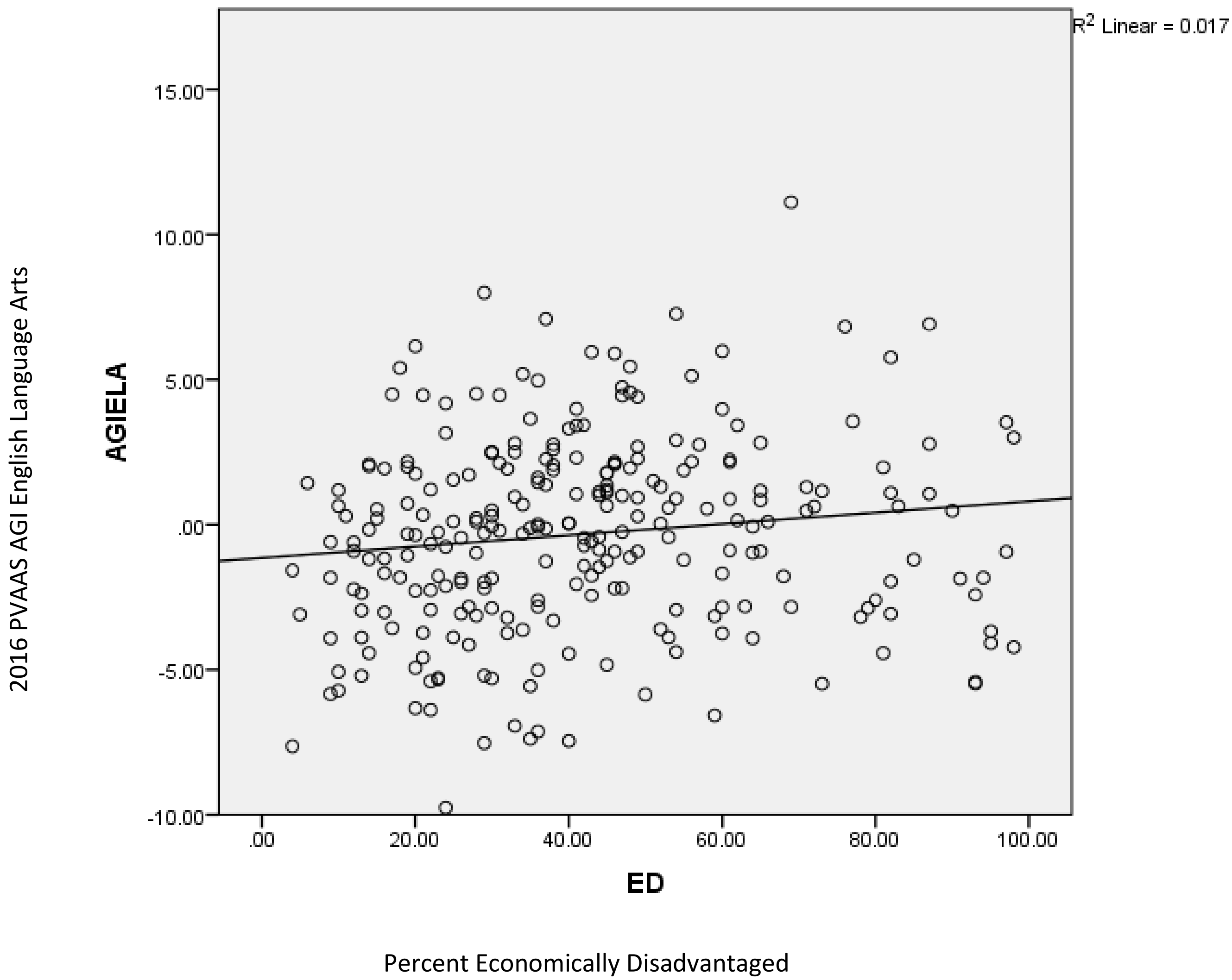

Figure 18. 2016 PVAAS AGI for English language arts related to percent economically disadvantaged.

There was no significant correlation between the 2016 PVAAS AGI English language arts score for each school building and the percentage of students requiring special education services in each building, r = -.06, n= 260, p = .168. Figure 19 represents the distribution of 2016 PVAAS AGI English language scores and the line of best fit. However, in this case the best fit line is not indicative of a relationship due to the lack of significance in the correlation.

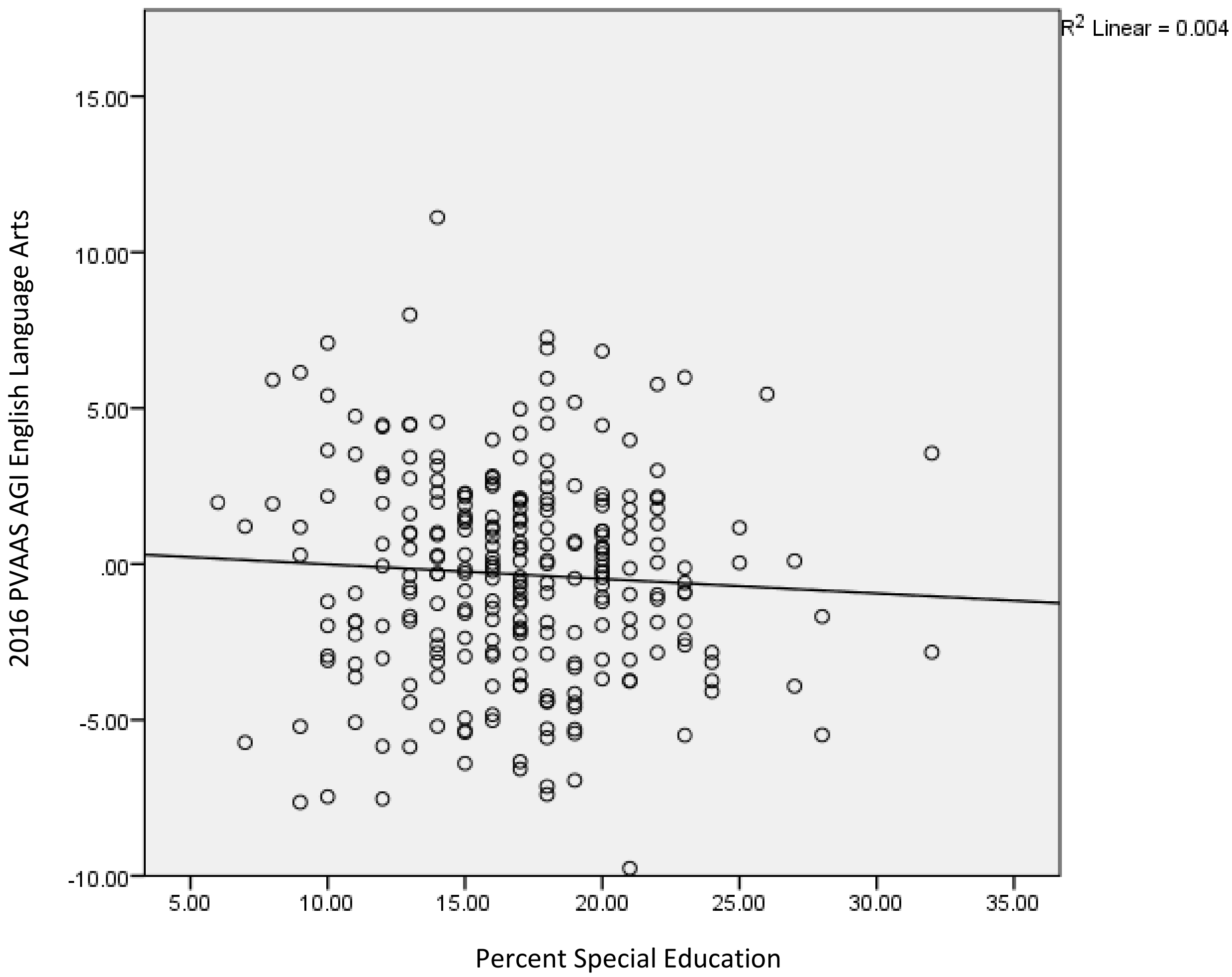

Figure 19. 2016 PVAAS AGI for English language arts related to percent of students needing special education services.

There was no significant correlation between the 2016 PVAAS AGI English language arts score for each school building and the percentage of students classified as limited English proficiency in each building, r = -.03, n= 260, p = .313. Figure 20 represents the distribution of 2016 PVAAS AGI English language arts scores and the line of best fit. However, in this case the best fit line is not indicative of a relationship due to the lack of significance in the correlation.

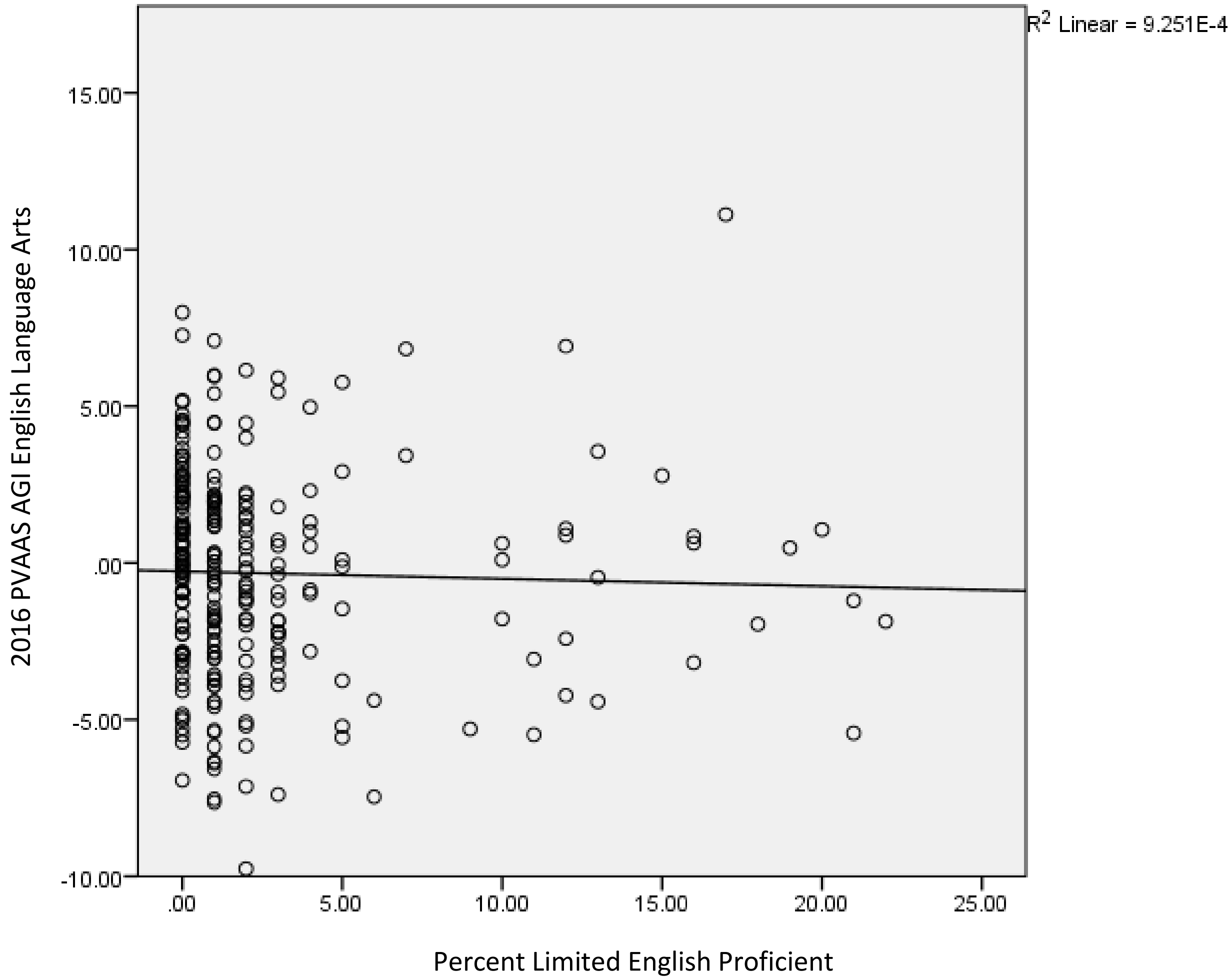

Figure 20. 2016 PVAAS AGI for English language arts related to percent limited English proficient.

There was no significant correlation between the 2016 PVAAS AGI English language arts score for each school building and the percentage of student classified as minorities in each building, r = -.09, n =260, p = .074. Figure 21 represents the distribution of 2016 PVAAS AGI English language arts scores and line of best fit. However, in this case the best fit line is not indicative of a relationship due to the lack of significance in the correlation.

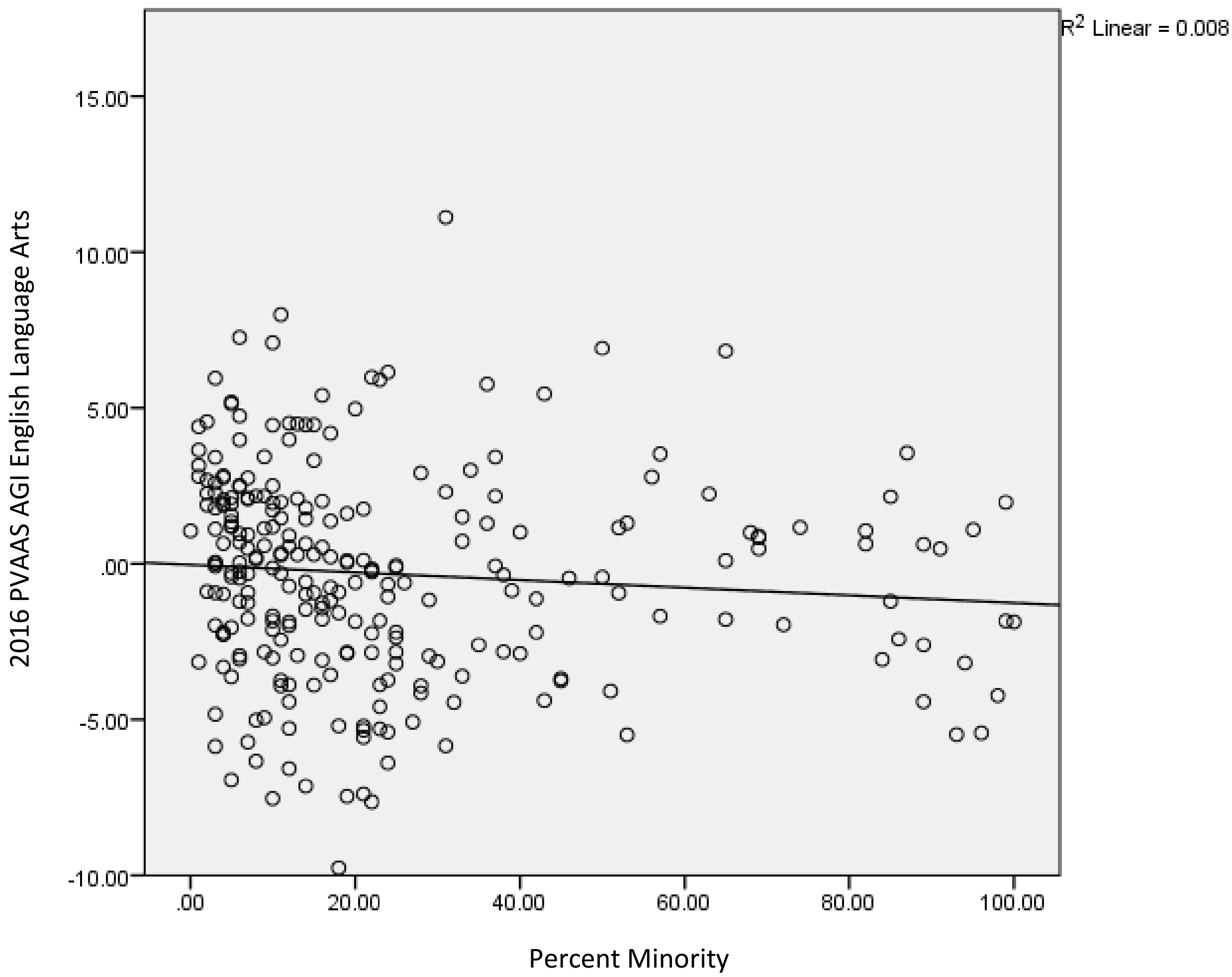

Figure 21. 2016 PVAAS AGI for English language arts related to percent minority.

Teacher Perceptions

Teachers receiving PVAAS AGI building and teacher scores for mathematics or English language arts were surveyed (Appendix E) at two middle schools servings grades six through eight within the same school district. The survey was distributed electronically to thirty five teachers. Twenty-two or 63% of the teachers responded to the survey. The first section of the survey requested demographic data. Teaching experience varied across those surveyed, with eight of the twenty-two (36.4 %) teachers reporting that they had at least fifteen years of teaching experience. Seven of the twenty-two (31.8%) teachers stated they had 11-15 years of teaching experience. Five of the twenty-two (22.7%) respondents indicated with six to ten years of experience and two of the twenty-two (9.1%) respondents had been teaching for at most five years (see Table 17).

Table 17

Years of teaching experience

Total years in teaching	Frequency	Percent
0-5	2	9.1
6-10	5	22.7
11-15	7	31.8
15 or more	8	36.4

Note. N=22.

Ten of the respondents claimed to be teachers of record for English language arts, indicating their PVAAS AGI score is derived from the English language arts scores for their school building, eleven of the respondents claimed to be teachers of record for mathematics, indicating their PVAAS AGI scores is derived from the mathematics scores for their school building, and one teacher chose not to indicate for which area they are a teacher of record (see Table 18).

Table 18

Subject area of record for which a PVAAS AGI score is received in the Classroom Rating Tool

Subject	Frequency	Percent
English language arts	10	45.5
Mathematics	11	50
Not indicated	1	4.5

Note. N = 22

One of the teachers selected did not to respond to the question indicating which subgroup of students they have experienced teaching. Of the 21 remaining responses, 100% indicated having experience teaching students who receive special education for learning disabilities, 20 (95.2%) indicated having experiences teaching economically disadvantaged, minority, and gifted students, and 17 (81%) indicated having experience teaching students with limited English proficiency (English as a second language) (see table 19).

Table 19

Experience teaching demographic subgroups

Demographic	Frequency	Percent
Special Education	21	100
Economically Disadvantaged	20	95.2
English as a Second Language	17	81
Minority	20	95.2
Gifted	20	95.2

Note. N=21.

Finally, twenty of the twenty-two respondents indicated that they had discussed the validity of the PVAAS statistical methodology for measuring teacher effectiveness with colleagues.

The second section of the survey utilized a likert scale to measure teacher perceptions of PVAAS AGI related issues that may emerge in school buildings and teachers' understanding and perceptions of PVAAS (see table 20). The purpose of the survey was to answer the following research hypotheses:

Research Hypothesis 1: Teachers believe that there is a shift in school morale as a result of their PVAAS AGI being included in annual effectiveness rating.

Research Hypothesis 2: Teachers experience a change in the instructional practices they are expected to utilize as a result of their school's PVAAS AGI.

Research Hypothesis 3: Teachers experience staffing changes as a result of their school's PVAAS AGI.

Research Hypothesis 4: Teachers experience scheduling changes as a result of their school's PVAAS AGI.

Research Hypothesis 5: Teachers perceive a shift in the amount of support provided to students as a result of changes in the PVAAS AGI.

Research Hypothesis 6: Teachers do not believe that their PVAAS AGI score is solely dependent on teacher efficacy.

Research Hypothesis 7: Teachers have an incomplete understanding of how the PVAAS AGI score is calculated.

Table 20

Teacher responses to survey statements using likert scale

Statement	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)	Mean
I believe there is a relationship between teacher morale and receiving a PVAAS Average Growth Index score in the Classroom Rating Tool.	1 (4.5%)	0 (0%)	2 (9.1%)	13 (59.1%)	6 (27.2%)	4.05
I believe that teacher morale has become significantly more positive as a result of teachers receiving a PVAAS Average Growth Index score in the Classroom Rating Tool.	10 (45.5%)	11 (50%)	1 (4.5%)	0 (0%)	0 (0%)	1.6
I believe that teacher morale has become significantly more negative as a result of teachers receiving a PVAAS Average Growth Index score in the Classroom Rating Tool.	0 (0%)	0 (0%)	0 (0%)	13 (59.1%)	9 (40.9%)	4.41
I believe that receiving a PVAAS Average Growth Index score in my Classroom Rating Tool has resulted in a change to the instructional practices I am expected to use in my classroom.	1 (4.5%)	3 (13.6%)	8 (36.4%)	8 (36.4%)	2 (9.1%)	3.32
Receiving a PVAAS Average Growth Index score in my Classroom Rating Tool has resulted in positive changes to the instructional practices that I am expected to use in my classroom.	3 (13.6%)	7 (31.8%)	10 (45.5%)	2 (9.1%)	0 (0%)	2.5
Receiving a PVAAS Average Growth Index score in my						

Classroom Rating Tool has resulted in negative changes to the instructional practices that I am expected to use in my classroom.	2 (9.1%)	6 (27.3%)	8 (36.4%)	4 (18.2%)	2 (9.1%)	2.91
Staffing changes have occurred as a result of a PVAAS Average Growth Index score being calculated for my school building.	0 (0%)	3 (13.5%)	15 (68.2%)	4 (18.2%)	0 (0%)	3.05
Scheduling changes have occurred as a result of a PVAAS Average Growth Index score being calculated for my school building.	0 (0%)	1 (4.5%)	5 (22.7%)	12 (54.5%)	4 (18.2%)	3.86
The amount of support all students receive in mathematics and reading has increased as a result of a PVAAS Average Growth Index score being calculated for my building.	1 (4.5%)	5 (22.7%)	3 (13.6%)	10 (45.5%)	3 (13.6%)	3.41
The amount of support all students receive in mathematics and reading has decreased as a result of the PVAAS Average Growth Index score being calculated for my building.	4 (18.2%)	12 (54.5%)	5 (22.7%)	0 (0%)	1 (4.5%)	2.18
The amount of support struggling students receive in mathematics and reading has increased as a result of a PVAAS Average Growth Index score being calculated for my building.	1 (4.5%)	5 (23.8%)	5 (23.8%)	9 (42.9%)	1 (4.8%)	3.05
The amount of support struggling students receive in mathematics and reading has decreased as a result of a PVAAS Average Growth Index score being calculated for my building.	2 (9.1%)	10 (45.5%)	7 (31.8%)	2 (9.1%)	1 (4.5%)	2.55
I believe that the PVAAS Average Growth Index score used in my						

Classroom Rating Tool is a valid measure of my effectiveness as a teacher.	10 (45.5%)	10 (45.5%)	1 (4.5%)	1 (4.5%)	0 (0%)	1.68
I believe that the PVAAS Average Growth Index score used in my Classroom Rating Tool is influenced by factors outside of my control.	0 (0%)	0 (0%)	1 (4.5%)	5 (22.7%)	16 (72.7%)	4.82
I understand how my PVAAS Average Growth Index score is calculated.	7 (31.8%)	10 (45.5%)	1 (4.5%)	4 (18.2%)	0 (0%)	2.1
I understand the statistical algorithm utilized to calculate my PVAAS Average Growth Index score.	14 (63.6%)	6 (27.3%)	1 (4.5%)	1 (4.5%)	0 (0%)	2.05

The first question in the likert scale survey asked teachers if they believed that there is a relationship between teacher morale and receiving a PVAAS Average Growth Index score in their classroom rating tool. Six (27.3%) respondents strongly agreed, 13 (59.1%) of respondents agreed, two (9.1%) were neutral, and one (4.5%) strongly disagreed. A teacher of English language arts with over 15 years of experience and indicated they strongly agreed with the statement explained: “As soon as you look at your score, you become depressed.” Another teacher of English language arts that indicated they agreed stated: “It’s hard to hear you need to do better or that we are striving for a certain color when you have worked really hard all year.” A teacher of English language arts with at most five years of experience indicating that they felt neutral stated: “Some teachers seem upset about it; others do not seem to mind. I am undecided. I do not know enough about it to make up my mind.” A teacher of mathematics with over fifteen years of experience offered the following insight: “The better your morale in any job, most likely the better you will perform.” Another teacher of mathematics with 11-15 years of teaching experience stated, “You want to feel validated for working hard. However, if the score is low it can really affect you negatively.” A teacher of mathematics with six to ten years of experience shared that: “It’s difficult not to take the rating personally when you know that you’ve worked so hard throughout the year to teach all of the skills. I’ve had both bad ratings and the highest rating, and work equally as hard to prepare my students.” Similarly, another teacher of mathematics with the eleven to fifteen years of experience believed: “In short a teacher can work very hard and still receive a “low” PVAAS score or vice versa. There are too many outside factors that affect students’ performance on the PSSAs and to tie that to a teacher is a very unfair practice. It lowers teacher morale.” A teacher of language arts with six to ten years of experience stated, “When your growth score is low you feel like administrators think you’re not doing a good job.” Finally, a teacher of mathematics with over 15 years of experience shared, “It is frustrating that teachers are partially evaluated on PVAAS scores, when educating a child is the responsibility not only the teacher, but the child, parent, and administration. It is not a one person job, and only math and LA teachers are affected. Other subject teachers do not have this pressure.”

Given that a significant number of explanations indicated that there is a relationship between PVAAS AGI scores and negative morale, it is not surprising that when asked to rate the statement “I believe that teacher morale has become significantly more positive as a result of teachers receiving a PVAAS Average Growth Index score in the Classroom Rating Tool, 11 (50%) selected disagree, ten (45.5%) selected agree, and one (4.5%) selected neutral. None of the respondents agreed or strongly agreed. However, ratings for the statement: “I believe that teacher morale has become significantly more negative as a result of teachers receiving a PVAAS Average Growth Index score in the Classroom Rating tool” indicated that 13 (59.1%) of responded agree and nine (40.9%) of responded strongly agree. Both of these questions were answered with explanations that illuminated only negative outcomes resulting from the use of PVAAS AGI scores in the Classroom Rating Tool. One teacher of mathematics, with 11-15 years of experience, stated: “In some cases I have personally seen it become where teachers won’t help other teachers, or share resources because they are so focused on increasing only their own scores.” Another teacher of mathematics with over 15 years of experience explained: “The testing window is a snapshot of the students’ progress, and there are so many other variables. There is a need for evaluating progress, but when so many changes are implemented each year, we can never figure out what works and what doesn’t. Plus, an educator can do all of the right instructional strategies and differentiate, but not all students will be star quarterbacks and not all students are motivated to learn. I wouldn’t blame my dentist for a cavity, if he taught me to floss, brush at least 3 times a day, and eat healthy if I didn’t follow his instructions.”

Teachers were then asked to rate statements on how receiving a PVAAS Average Growth Index score on the Classroom Rating Tool impacted instructional practices. Eight (36.4%) respondents agreed that that receiving a PVAAS Average Growth Index score on their Classroom Rating Tool resulted in a change to the instruction practices they are expected to use in the classroom, eight (36.4%) felt neutral about this statement, three (13.6%) respondents disagreed, and one (9.1%) strongly disagreed. One teacher, presumably a teacher of special education and a teacher of record for English language arts stated: “I teach my students according to IEP goals; this gives me an abundance of concepts to teach. Sometimes I am given far too much data on my students-they have many gaps, and I try to pick the most important gaps to fill.” One respondent, a mathematics teacher with 11-15 years of experience, stated: “It’s not the PVAAS score that has changed my instructional practices, but the content on the PSSA itself." Other responses included, "Still work just as hard", "the changes are not positive", and "It makes you feel like you need to teach to the test due to the added pressure".

When teachers were prompted to rate the statement, "Receiving a PVAAS Average Growth Index score in my Classroom Rating Tool has resulted in positive changes to the instructional practices that I am expected to use in my classroom, three (13.6%) strongly disagreed, seven (31.8%) disagreed, 10 (45.5%) were neutral, and two (9.1%) agreed. None of the respondents strongly agreed. When prompted to rate the statement "Receiving a PVAAS Average Growth Index score in my Classroom Rating Tool has resulted in negative changes to the instructional practices that I am expected to use in my classroom, two (9.1%) responded that they strongly disagree, six (27.3%) disagreed, eight (36.4%) were neutral, four (18.2%) agreed, and two (9.1%) strongly agreed. A teacher of English language arts with 11-15 years of experience explained their reasoning for disagreeing with the change being positive, "Sure I have learned more strategies but I have had to also get rid of other effective techniques due to time." A teacher of mathematics with 11-15 years of experience that strongly disagreed with the statement that changes were positive explained: "From my perspective, I think it has made me a more nervous teacher. It is taking away the joy of teaching." The same teacher also stated: "PVAAS scores have created an environment of teach, teach, teach to the standards on the test, instead of focusing on the joy of learning mathematics." One teacher that indicated neutral in their response to both questions stated: "I feel like the changes we are all making are taking the "fun" of math out of our practices. Students are having extra math/reading instruction which is making them sick of the subjects… "overkill". They also stated: "I use a lot more formative assessment than ever, I do have a better idea of what they can/can't do." Other responses included: "I have done well, so I feel okay about it; however, I do see the other side" and "I have made changes that I feel are helping my students take standardized assessments, but I feel as though I and now teaching to the test”.

When asked to rate the statement “Staffing changes have occurred as a result of a PVAAS Average Growth Index score being calculated for my school building”, 15 (68.2%) responded neutral, four (18.2%) responded that they agree, and three (13.6%) responded that they disagree. No respondents indicated that they strongly agree or strongly disagree. One respondent stated their reasoning for agreeing with the statement: “Extra reading/math in 6th grade…no more FCS which they loved.” Another respondent stated: “I am not certain, but I do know that myself and a few colleagues believe that staff changes are a result of scores.” Other explanations indicated that the respondents were unsure if there was any relationship between staffing changes and the PVAAS Average Growth Index score calculated for each school building.

The majority of teachers either agreed or strongly agreed with the statement: “The amount of support all students receive in mathematics and reading has increased as a result of a PVAAS Average Growth Index score being calculated for my building. Three (13.6%) respondents strongly agreed, 10 (45.5%) respondents agreed, three (13.6%) indicated neutral, five (22.7%) disagreed, and one respondent (4.5%) strongly disagreed. One respondent, a teacher of mathematics with over 15 years of experience, who strongly agreed that all students receive more support stated: “The kids are hammered with both math and LA, and they get burned out”. Another mathematics teacher that had indicated they agree that all students receive more support referenced: “extra explorations classes created”. One teacher of English language arts indicated that they believe students receive more support “in reading more than math”. A teacher responding that they disagree with the statement that all students receive less support explained, "6th grade has an extra math and language arts class every other day although there is no specific or set curriculum. All students have longer math and language arts classes-increased from 45-53 minutes."

When asked about the relationship between the amount of support struggling students receive and PVAAS Average Growth Index scores being calculated for buildings, nine (42.9%) respondents agreed that: "The amount of support struggling students receive in mathematics and reading has increased as a result of a PVAAS Average Growth Index score being calculated for my building". Five (23.8%) respondents indicated neutral, one (4.5%) indicated that they strongly disagree and one (4.5%) respondent indicated that they strongly agree. A teacher of mathematics with six to ten years of experience explained their neutral response: "Learning support math has own class now, but not much time for student to see teachers for help as when we had 10th period REM."

The majority of respondents disagreed or strongly disagreed with the statement: "The amount of support struggling students receive in mathematics and reading had decreased as a result of a PVAAS Growth Index score being calculated for my building". Two (9.1%) respondents strongly disagreed, ten (45.5%) disagreed, seven (31.8%) felt neutral, two (9.1%) agreed and one (4.5%) strongly agreed. A teacher of English language arts with 11-15 years of experience explained why they disagreed with this statement, "It's ironic how we have less kids in reading support now than we did before PVAAS."

In response to the statement "I believe that the PVAAS Average Growth Index scores used in my Classroom Rating Tool is a valid measure of my effectiveness as a teacher and overwhelming majority of respondents indicated that they either strongly disagreed or disagreed.

10 (45.5%) strongly disagreed, 10 (45.5%) disagreed, one (4.5%), felt neutral, one (4.5%) agreed with the aforementioned statement. None of the respondents strongly agreed with the statement. An English language arts teachers with more than 15 years of experience explained why they strongly disagreed, “The state can’t even explain how they get the score. Ridiculous.” Another English language arts teacher with 11-15 years of experience explain why they strongly disagreed as well, “…because there are other factors on student scores that weigh more than my teaching…” A mathematics teacher with over fifteen years of experience explained why they strongly disagreed: “It is affected by student ability and performance which can greatly vary from year to year.” A mathematics teacher with six to ten year of teaching experience explained why they disagreed with this statement: “If my students don’t care about how they do on the test, how can that show MY effectiveness? Just seems like a “mystery test” (why we can’t view an old version of it to see what it is like, besides the released items), mystery formula for scoring….? Ugh.”

Sixteen (72.7%) of the 22 respondents strongly agreed with the statement “I believe that the PVAAS Average Growth Index score used in my Classroom Rating Tool is influenced by factors outside of my control, while five (22.5%) agreed with the statement, and one (4.5%) felt neutral about the statement. None of the respondents disagreed or strongly disagreed. Three respondents explained why they strongly agreed with the statement. An English language arts teacher with 11-15 years of experience described other factors that they believed may influence the PVAAS Average Growth Index score: “student motivation, sleep, food, economics, emotional stability, whether academics are valued at home”. A mathematics teacher with over 15 years of experience stated: “It is affected by student ability and performance, which can greatly vary from year to year.” Another mathematics teacher with six to ten years of experience explained: “I feel that support at home, a student’s grit/work ethic, etc. also play a big part in this.”

Finally, respondents were prompted to rate two statements designed to determine how well they understand the PVAAS Average Growth Index score. When responding to the statement: “I understand how my PVAAS Average Growth Index score is calculated, seven (31.8%) respondents strongly disagreed, ten (45.5%) respondents disagreed, one (4.5%) respondent felt neutral, and four (18.2%) respondents agreed. None of the respondents strongly agreed. When responding to the statement “I understand the statistical algorithm utilized to calculate my PVAAS Average Growth Index score, 14 (63.3%) respondents strongly disagreed, six (27.3%) disagreed, one (4.5%) respondent felt neutral, and one (4.5%) agreed. None of the respondent strongly agreed with the statement. None of the respondents provided any explanation for their responses to either of these questions.

Overall, the response scores of English language arts teachers and mathematics teachers paralleled each other, with the exception of question six (see Figure 21). English language arts teachers seemed to find less of a relationship between PVAAS AGI scores and changes to instructional practices than mathematics teachers.

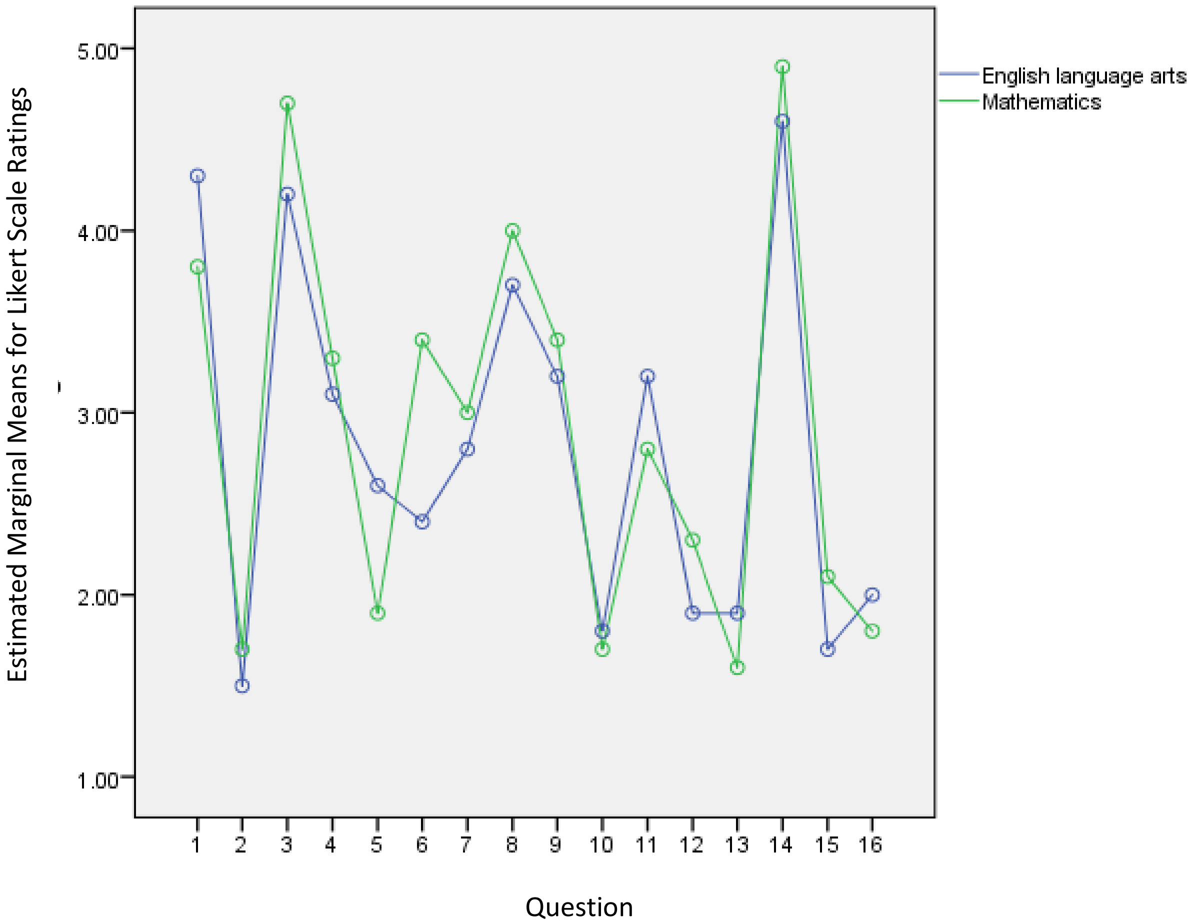

Figure 22. Relationship between responses of teacher of English language arts and Mathematics.

This phenomenon may be related to the fact that although both schools had poor PVAAS AGI scores, the percent proficient scores for English language arts was significantly higher for both schools that the percent proficient for mathematics. Having a high percent proficient may combat concerns about instructional practices. Similarly, having a low percent proficient may prompt changes. Table 21 represents the PVAAS AGI and percent proficient scores in English language arts and mathematics in each school.

Table 21

PVAAS AGI and percent proficient for school in which teachers were surveyed

Subject	School	PVAAS AGI	Percent Proficient
English language arts	School A	-3.18	82.8
English language arts	School B	-2.86	79.2
Mathematics	School A	-2.61	55.3
Mathematics	School B	-2.88	50.8

Summary

The results of the research accepted the null hypotheses that there would be no difference in mean gain scores of schools categorized by average normal curve equivalent (NCE) scores for both mathematics and English language arts. However, the research did indicate a significant negative change in PVAAS AGI scores over time in both mathematics and English language arts across all performance categories as well as inaccuracies in public reporting of school data by PVAAS. The mean gain scores for schools at all performance levels fell from 2013 to 2014 and from 2015 to 2016. A couple of performance level categories experienced a slight and brief increase from 2014 to 2015 which may correlate to a change in the standards addressed by the PSSA test in the 2015 school year.

Very weak or no correlations were detected between demographic categories and PVAAS AGI in mathematics and English language arts for both 2015 and 2016. Individual Pearson correlation tests between the percent of the student population in each school occupied by each demographic category and the PVAAS AGI score for mathematics and English language arts were calculated. In 2015, these calculations indicated very weak negative correlations between PVAAS AGI in mathematics and the percentage of economically disadvantaged students, special education students, limited English proficient, and minority students. A very weak positive correlation was detected between PVAAS AGI in English language arts and the percentage of economically disadvantaged students in a school in 2015. Very weak negative correlations between PVAAS AGI in English language arts and the percentage of special education, and minority students were identified in 2015. No significant correlation between PVAAS AGI in English language arts and limited English proficiency students was identified in 2016. In 2016, no significant correlations were identified between PVAAS AGI in mathematics and percent economically disadvantaged or percent of special education students. Very weak correlations were identified between PVAAS AGI in mathematics and the percentage of English language learners and minorities. A very weak correlation was identified between PVAAS AGI for English language arts and the percent of economically disadvantaged students, however no correlations was found between PVAAS AGI for English language arts and the percentages of special education students, English language learners, and minorities in 2016.

The majority of teachers indicated negative effects resulting from PVAAS AGI scores being reported in the classroom rating tool. The inclusion of PVAAS AGI scores in the Classroom Rating Tool was associated with negative morale, scheduling changes, and changes in student support provided. Teachers also tended to indicate that they do not understand whether or not the calculation is valid and how PVAAS AGI scores are determined.

The results of this analysis yielded interesting information regarding the use of PVAAS AGI to measure teacher performance. The analysis first demonstrated that there is a deficiency in the accuracy of public reporting and possibly in the accuracy of the calculations themselves, as a number of schools were found to be incorrectly labeled by grade level. Secondly, the fact that the

PVAAS AGI for schools at all performance levels fell significantly in both English language arts and mathematics raises questions about the calculation, as well as the standards being measured and test design. Despite strong correlations between average NCE scores and the percentage of economically disadvantaged across years and subjects, the demographic variables tested do not seem to influence the school level PVAAS AGI score. Finally, the analysis made it evident that teacher morale is tied to school performance and if PVAAS AGI is part of the school performance equation, it influences teacher morale. Additionally, it is possible that teachers believe that schools may make scheduling and resource decisions based on PVAAS AGI. Both of these findings demonstrate the importance of receiving an accurate measure of school and teacher performance. Finally, the majority of teachers surveyed indicated that they do not know if their PVAAS AGI score is valid and they do not understand the statistical algorithm that is used to calculate PVAAS AGI. This raises concerns about transparency in performance evaluations.

Chapter 5: Conclusions and Recommendations

In 2013 the Pennsylvania Department of Education began using a statistical algorithm, called a Value Added Model (VAM), measuring growth in percentile rankings of students on standardized tests from year to year, for the purpose of determining the efficacy of teachers, schools, and school districts. This growth is determined by the Pennsylvania Value-Added Assessment System (PVAAS) and the growth measure for teachers, schools, and school districts is referred to as the Average Growth Index (AGI). Act 82 of 2012 ushered in the Classroom Rating Tool, a formula for calculating an annual evaluation score for each public school teacher in Pennsylvania. For teachers of mathematics and English language arts in grades six through eight, 15% of the overall score is comprised of the PVAAS AGI for all of a teacher's students and 15% of the overall score is comprised of the PVAAS AGI for all students in the school. Teachers now rely on the validity of PVAAS's methodology for determining AGI scores to receive accurate annual evaluations.

Education researchers have questioned the validity of utilizing a VAM to measure teacher efficacy. Concerns include the possibility of a ceiling effect, an inability for students with high normal curve equivalent scores to demonstrate as much growth as students with lower normal curve equivalent scores, external factors such as demographic variables influencing AGI, and the impact this system of measurement may have on teachers, students, and decision making in schools. The American Statistical Association claims that VAMS should not be used to evaluate teacher efficacy because they can only identify correlations and not causation (ASA, 2014). Despite this recommendation, Pennsylvania, like many other states seeking to meet the requirements of Race to the Top (RTTT) and the Elementary and Secondary Education Act

(ESEA) Waiver, developed and implemented a system for evaluating teachers using VAMs. This system was entered into Pennsylvania state education law in 2012 as Act 82.

The purpose of this research was to determine if the Pennsylvania Value-Added Assessment System Average Growth Index (PVAAS AGI) scores for schools provides a valid and consistent assessment of teacher effectiveness and if teacher perceptions indicate that the validity and consistency of the PVAAS AGI scores included in the Classroom Rating Tool are important. Data collection and analysis were guided by three research questions:

1. To what extent do PVAAS AGI scores assigned to a school change statistically over time?
2. How does the PVAAS AGI scores assigned to a school relate to the percentage of economically disadvantaged, learning disabled, English language learners, and minorities in the school?
3. To what extent do teacher perceptions of the impact of PVAAS AGI indicate that the validity of the PVAAS AGI calculated for their school is important?

A mixed ANOVA model was utilized to identify differences in mean gains in PVAAS AGI for mathematics and English language arts between 260 schools serving grades six through eight with differing achievement levels. The 260 schools were categorized by achievement levels based on the average Normal Curve Equivalent (NCE) score of their students on the Pennsylvania Standardized System of Assessment (PSSA). Schools were classified using average NCE score data from the base year, 2013, and then a mixed ANOVA was performed using four years of PVAAS AGI data for each school. The mixed ANOVA was performed once using data from the English language arts PSSA and once using data from the mathematics

PSSA. Pearson correlation tests were conducted using 2015 and 2016 PVAAS AGI data from the same schools in mathematics and English language arts and the percentages of economically disadvantaged, special education, limited English proficient and minority students in each school. A Pearson correlation test was also conducted between the average NCE scores in English language arts and mathematics for each school and the percentages of economically disadvantaged, special education, limited English proficient, and minority students in each school. A likert scale survey was distributed to 35 PVAAS eligible teachers of grades six through eight working across two middle schools serving grades six through eight within the same district. The survey was distributed via Google forms. Twenty-two of the 35 teachers responded to the survey.

Conclusions

PVAAS AGI Statistical Changes over Time

The research indicates that a ceiling effect did not influence the PVAAS AGI scores for middle schools serving grades six through eight in mathematics or English language arts from 2013 to 2016 in this study. There were no correlations between achievement level of each school and the mean PVAAS AGI gains in both mathematics and English language arts. The mean gains in PVAAS AGI for both subjects fell significantly for schools over four years of testing data. Given that the purpose of utilizing this system is to promote student growth, these results raise questions surrounding not only the statistical algorithm itself, but the data it incorporates, and the design of the tests being utilized. It is important to note that contents of the test changed in 2015. However, the mean gain scores in mathematics and English language arts fell both before the change in the test and after the change in the test. One would expect the dip in the mean gains to occur in the year that new testing content was implemented, as schools adjust to the change. Instead, schools in some achievement levels experienced an increase in the mean gains during the new test year and then significantly fell in the following year.

It is interesting to note that in both the case of mathematics and English language arts, the achievement levels that experienced an increase in mean gain scores in 2015 were the two highest achieving levels, level 1 and level 2, and the lowest achieving level, level 5. While there was not a statistically significant number of schools categorized as level 5 in mathematics (n=3) and English language arts (n=1), the increase in the two highest levels, which did contain a significant number of data points, raises questions as to why these groups were able to show positive gains on the new test, while lower achieving groups demonstrated negative gains.

Another question raised by the overall decrease in mean gains of PVAAS AGI scores, is whether an overall decrease should even be statistically possible. Given that PVAAS AGI is calculated using students' Normal Curve Equivalent (NCE) scores, one might expect that the mean gain for all schools would be zero. This is because NCEs are a system of ranking from one to 99. In the simplest of terms, as one student gains in ranking position another student must fall. If all students gain equally from one year to the next, each student's rank remains the same, growth is zero (indicating one year's growth), and each rank indicates a higher level of achievement (Area Education Agency, 2012). The SAS Institute explains their reasoning for using NCEs in a white paper titled SAS EVAAS K-12 Statistical Models (SAS Institute, 2016): "…NCEs have a critical advantage for measuring growth: they 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. NCEs are constructed to be equivalent to percentile ranks at 1, 50, and 99, with the mean being 50 and the standard deviation being 21.063 by definition.

Although percentile ranks are usually truncated below 1 and above 99, NCEs are allowed to range below zero and above 100 to preserve their equal-interval property and to avoid truncating the test scale. Truncating would create an artificial ceiling or floor, and this could bias the results of the value-added measure for certain types of students by forcing the gain to be close to zero or even negative" (SAS Institute, 2016, p.5). Despite stating that the reason for using NCEs is because it is possible to avoid truncating scores, the document goes on to describe the process that EVAAS utilizes as a process that truncates scores. "…NCEs are scaled so that they exactly match 1, 50, and 99" (SAS Institute, 2016, p.7). According to this statement, EVAAS uses NCEs in such a way that forces a mean of zero or a negative mean. Given that the mean gains fell for all middle schools serving grades six through eight analyzed in this study, one could either assume that this is a result of how EVAAS uses NCEs, causing forced negative gains. Alternatively, if the mean gains are forced to zero, then equal positive gains would be expected among the populations of students in six, seventh, and eight grade that attend schools serving other grade levels. If the latter is the case, this may indicate that the middle school model serving grades six through eight is less effective than schools using different models, such as schools serving grades five to eight, or schools that break down the middle school years in five through six and seven through eight. The use of NCEs in such a way that forces a negative mean or a mean of zero raises ethical concerns regarding the use of a calculation that requires negatives gains to equal positive gains to evaluate teacher efficacy.

During the data collection process a number of discrepancies were identified on the state data reporting site for the Pennsylvania Value Added Assessment System (SAS Institute Inc., 2016). Eleven percent of the schools reviewed in the data collection process were found to be incorrectly identified by grade on the website. There were also several cases in which the data for specific schools could not be found for specific years on this data reporting site.

Correlations between PVAAS AGI and Demographic Variables

PVAAS AGI is calculated using the average NCE score of each school. Therefore, before testing for correlations between PVAAS AGI and demographic variables, correlations between the average NCE and demographics were first determined to assist in a thorough understanding of the results. A strong negative correlation was found to exist between the average NCE score for a school and the percentage of economically disadvantaged students in the school for mathematics and English language arts in 2015 and 2016. A moderate negative correlation between average NCE scores and the percentage of minority students was also identified in both mathematics and English language arts in 2015 and 2016. While the correlation between the percentage of special education students and average NCE in mathematics and English language arts was found to be weak negative in 2015, a moderate negative correlation was identified between these variables in 2016. A weak negative correlation was identified between average NCE scores and limited English proficiency in both mathematics and English language arts in 2015 and 2016.

Correlations between PVAAS AGI and demographics were then identified. The correlations found between the percentage of students classified as economically disadvantaged, special education, limited English proficient, and minority were found to be very weak or not existing. Few of the very weak correlations are consistent across 2015 and 2016. Although a very weak correlation existed between PVAAS AGI in mathematics and percent economically disadvantaged in 2015, no correlation between these two variables existed in 2016. Similarly, the very weak correlations between PVAAS AGI in mathematics and English language arts and the percentage of special education students in 2015 did not exist in 2016. The very weak correlation identified between PVAAS AGI in English language arts and minorities in 2015 did not exist in 2016.

There were a few consistencies in correlations across years. A very weak correlation detected between PVAAS AGI in mathematics and limited English proficiency was consistent from 2015 to 2016. Similarly, a very weak correlation between PVAAS AGI in mathematics and minorities was consistent from 2015 to 2016. However, no correlation between PVAAS AGI in English language arts and limited English proficiency was detected in both 2015 and 2016.

In general, the analysis indicates that the four demographic variables reported by the state do not seem to influence the PVAAS AGI for mathematics and English language arts assigned to school buildings, despite the fact that there were strong negative correlations between average NCE scores and the percentage of economically disadvantaged students. This finding supports the Pennsylvania Department of Education's claims that the value-added model utilized to calculate PVAAS AGI does not need to account for demographics, because they are intrinsically accounted for by measuring a student against their own data from previous years (EVAAS, 2015).

Teacher Perceptions of the PVAAS AGI

Analysis of the survey concluded that the majority of the teachers believed that there is a shift in school morale as a result of their PVAAS AGI being included in their annual effectiveness rating. The majority of these teachers found the shift in morale to be negative, citing direct connections to the PVAAS AGI score. For example, one teacher stated: "As soon as you see your score, you become depressed." This negative morale may correlate to the fact that both schools received negative PVAAS AGI scores. On average, mathematics teachers indicated that morale was more negative than teachers of English language arts. This may be attributed to the fact that although both received negative PVAAS AGI scores, English language arts teachers had a higher number of students scoring proficient than mathematics teachers. This indicates that other data may also influence teacher morale either by mediating positive and negative feelings or exacerbating these feelings.

The majority of teachers felt neutral or agreed that there has been a change to expected instructional practices as a result of their school's PVAAS AGI. However, when asked if this change was positive or negative, the results were mixed, with the majority of teachers selecting neutral or disagreeing with both. This indicates that teachers perceive some changes but they are not sure yet whether these changes will have positive outcomes.

The majority of teachers indicated that they were unsure if staffing changes occurred as a result of a PVAAS Average Growth Index. Although the majority of teachers indicated neutral, their explanations indicated that they were suspicious of staffing changes resulting from PVAAS AGI scores, but unsure. One respondent stated: "I am not certain, but I do know that myself and a few colleagues believe that staff changes are a result of scores."

The majority of teachers indicated that they believe scheduling changes occur as a result of their school's PVAAS AGI. This is a significant finding as it indicates that important decisions are possibly being made within schools based on PVAAS AGI. If PVAAS AGI is not a valid measure of student growth, this would mean that decisions are being made based on invalid data.

Teachers indicated that receiving a PVAAS AGI score does not relate to the amount of support received by all students. The majority of teachers indicated agreement with the statement that the amount of support all students receive in mathematics and English language arts and disagreement with a statement indicating that the amount of support provided to all students decreased. When prompted to respond to statements regarding an increase or decrease of support for struggling students related to PVAAS AGI, the majority of teachers indicated that they agreed that support for struggling students had increased and disagreed with the statement that support for struggling students had decreased.

Sixteen of the twenty-two respondents indicated that they strongly agree that their PVAAS AGI score is influenced by factors outside of their control. Five of the twenty-two respondents indicated that they agree that their PVAAS AGI scores are influenced by factors outside of their control. Only one respondent indicated that they had neutral feelings about this statement and none of the respondents disagreed with the statement.

The results of the analysis indicated that the majority of teachers do not believe that their PVAAS AGI score is solely dependent on teacher efficacy. The majority of teachers also indicated that they do not understand the statistical algorithm that is used to calculate PVAAS AGI or, more generally, how the PVAAS AGI score is calculated.

Recommendations

Actionable Solutions

Given that this research indicated negative mean gains at all achievement levels, it is not clear that PVAAS AGI data can be relied upon to make decisions that impact students academically. Therefore, it is recommended that until further research validates the use of

PVAAS AGI scores for decision making; that schools continue to rely solely on the data from proficiency levels provided by the Pennsylvania Department of Education and also reported on the PVAAS website for decision making purposes. Scheduling and staffing changes should be based on needs indicated by proficiency levels as opposed to PVAAS AGI. Similarly, changes to instructional practices should be based on research in best practices, as well as, the proficiency levels of students.

Additionally, it is recommended that the discrepancies in reporting school grade levels on the state data reporting site (SAS Institute Inc., 2016) be remedied for past, current, and future years. Since the school performance levels reported on this site are used for school-based decision making, it is important for the Pennsylvania Department of Education (PDE) to develop a system for ensuring that all data reported on this public website is accurate. It is also recommended that PDE ensure that the grade level reporting discrepancies identified do not impact the PVAAS AGI scores for these buildings. It is unknown whether grade level data from other schools was incorporated in the PVAAS AGI calculation for schools whose grade levels were reported incorrectly.

There is a need for further research to determine the correlation between PVAAS AGI scores for individual teachers and teacher observation scores utilizing data from criterion-referenced assessments. There is existing research demonstrating that this methodology does correlate to observation scores when using scale scores from norm-referenced assessment items (Sanders & Horn, 1994). Therefore, it is recommended that norm-referenced assessment items be included within each PSSA test and that PVAAS AGI scores be calculated using scale scores from only the norm-referenced assessment items until further research demonstrates that there is a similar positive correlation between a teacher's PVAAS AGI and their observation scores when using criterion referenced tests.

PVAAS AGI scores have caused a negative impact to morale for teachers who participated in the study. It is possible that the impact on morale is negative because the schools in which teachers were interviewed received negative PVAAS AGI scores. Given that the mean gains in PVAAS AGI decreased for all 260 schools analyzed in this study, the small sample utilized in this survey study may be indicative of a more wide spread problem of negative teacher morale resulting from PVAAS AGI. Until it is concluded definitively that PVAAS AGI is a valid measure of teacher performance, down playing these scores and increasing focus how students are visibly succeeding in school may improve teacher morale.

The survey also indicated that teachers do not trust that the PVAAS AGI score is not influenced by outside factors nor do they have a complete understanding of how it is calculated. Given that the PVAAS AGI for schools and individual teachers is included in the annual evaluation for each teacher, it is important that teachers understand how their score is determined and how they can improve. Increasing professional development for teachers on how the Pennsylvania Value-Added Assessment System measures student growth and calculates AGI scores for buildings and teachers can support teachers in having a better understanding of their annual evaluations provided using the Classroom Rating Tool.

The need for further research to validate PVAAS AGI as an accurate method for measuring teacher performance makes it critical that schools do not use these scores to make decisions regarding employment and tenure. This could become a concern because the PVAAS AGI scores are provided on teachers' annual evaluations in the Classroom Rating Tool. These evaluations provide documentation regarding teacher's performance and determine whether a teacher is rated Satisfactory or Unsatisfactory for the year (see Appendix A). It is recommended that schools forego requesting the Classroom Rating Tool of experienced teachers applying for employment. Similarly, it is recommended that these scores not be used to terminate employment or initiate disciplinary action.

Recommendations for Further Research

Further investigation is needed to develop a better understanding of the accuracy of the PVAAS AGI calculation. First, it is necessary to understand whether Normal Curve Equivalent scores are truncated before use in the value added model. A study of all students in Pennsylvania by individual grade level may illuminate whether a particular middle school model has higher outcomes than other models. This will also help us understand if there are particular grade levels that tend to experience greater difficulty on the Pennsylvania State System of Assessment (PSSA) test in mathematics or English language arts. Alternatively, this research may bring to light problems with the model itself or how the model is applied.

One of the issues impacting the fifteen percent of the Classroom Rating Tool occupied by the individual teacher's PVAAS AGI and not addressed by this research, is the fact that schools may not use random sampling when distributing students among teachers. This researcher's personal experience contributes to this understanding, as in all of the school districts this researcher has taught, students receiving special education services were placed in the same class together so that one paraprofessional could support them as they go from class to class. Distributing students with increased needs evenly across classrooms is sometimes difficult for school districts, as there are not enough financial resources to provide the appropriate supports.

The lack of random sampling may have a greater impact on individual teacher's PVAAS AGI results, particularly when dealing with smaller data sets. Research on how students are assigned to given teachers and identifying correlations between individual teacher PVAAS AGI scores and the percentage of students in a given demographic subgroup, can shed some light on the extent to which the lack of random sampling impacts individual teacher PVAAS AGI scores.

The literature review surfaced several other issues that were not directly addressed in this study, however beg additional research. Research indicates that children's brains do not develop at the same rate nor is that rate linear, which could potentially skew the results derived by a model that inherently expects a linear growth rate (UNICEF, n.d.). Similarly, second language learners acquire the English language at different rates depending on their age, thereby possibly causing a disparity in growth that is not attributable to the teacher's efficacy (Phillips, 2002). Other factors that are not attributable to a teacher's efficacy, but that were not addressed in this research study are the effect of class size, the frequency of support services provided to students, and the ratio of support staff to students. Finally, of greater concern is a contradiction found in the descriptions PVAAS and PSSA testing characteristics. While the Data Recognition Corporation (DRC), the company that designs the PSSA, warns that scores at the maximum and the minimum end of the scale may not be accurate (DRC 2010, p. 232, para. 2) the SAS Institute claims that the tests utilized by their model must "…adequately measure the performance of both very low and very high achieving students" (Rivers, J.C., Sanders, W.L., Wright, J.T., & White, S.P., 2010, p.2). This begs the overarching question of whether it is even possible for the calculations produced by PVAAS to be accurate.

Summary

This research study did not conclusively determine that PVAAS AGI scores are influenced by the achievement level of students. The results did indicate that the mean gains for the 260 middle schools serving grades six through eight, regardless of achievement level, fell significantly over a four year period. Since the PVAAS AGI calculation uses normal curve equivalent (NCE) scores, a system of scoring that ranks students using equal intervals, to determine growth from one year to the next, questions remain regarding how the NCE scores are treated and if the method of treating these scores within the PVAAS statistical model forces the mean to zero, thereby invalidating the scores. However, if this is the case, these scores may demonstrate that students in grades six through eight, who attend schools with different grade level models that were not included in this study, produce higher gains, countering the negative gains produced in this study. Further research to determine how NCE scores are treated in the PVAAS model and how students in particular grade levels perform across all school models in Pennsylvania will provide greater context to this result of the study.

An evaluation of correlations between the four demographic categories reported by the Pennsylvania Department of Education for each school and PVAAS AGI scores from two consecutive years yielded very weak correlations or no correlation. Therefore, the percentage of economically disadvantaged students, special education students, English language learners, and minority students did not influence the PVAAS AGI scores in mathematics or English language arts for the 260 middle schools serving grades six through eight included in this study. There were strong negative correlations between the percentage of economically disadvantaged students and PVAAS AGI scores and moderately negative correlations between the percentage of minority students and PVAAS AGI scores. This indicates that the PVAAS AGI model utilized for calculating school scores does negate the effects of these demographic variables.

This analysis also discovered that morale was impacted negatively by PVAAS AGI scores in schools that have negative PVAAS AGI scores in mathematics and English language arts. It was found that teachers believe their school has made decisions about instructional practices implements and scheduling based on PVAAS AGI scores leading to the conclusion that schools make important decisions that rely on the validity of PVAAS AGI scores. Teachers also indicated that they do not understand how PVAAS AGI is calculated, leading to the conclusion that most teachers also do not understand how the PVAAS AGI portion of their annual evaluation, which comprises 30% of how their overall score is determined.

Finally, the process of this investigation illuminated a couple matters of interest. First, it was found that there were multiple discrepancies in the reporting of school grade levels on the designated state score reporting website (SAS Institute, 2016). A careful review of how school data is collected and reported will help ensure the accuracy of data reporting. Secondly, a statement published by the Data Recognition Corporation (DRC), regarding the fact that scores at the maximum and minimum ends of the scale may not be accurate (DRC 2010, p. 232, para. 2), conflicts with a statement published by the SAS Institute reporting that their statistical model requires that the testing data utilized in their model must adequately measure students who perform at the extreme ends of the scale (Rivers, J.C., Sanders, W.L., Wright, J.T., & White, S.P., 2010, p.2). These statements raise questions regarding whether the PVAAS models can accurately measure growth when utilizing data from tests developed by the DRC.

References

Althauser, K. L. (2010). *The effects of a sustained, job-embedded professional development on elementary teachers' math teaching self-efficacy and the resulting effects on their students' achievement.* Retrieved from http://www.library.drexel.edu/cgi-bin/r.cgi/login?url=http://search.proquest.com/docview/968114647?accountid=10559

American Statistical Association. (2014). *ASA statement on using value-added models for educational assessment*. Retrieved from https://www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf

Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, *37*(2), 65-75. Retrieved from http://search.proquest.com

Area Education Agency 267. (2012). *What is a normal curve equivalent score*? Retrieved from: https://www.aea267.k12.ia.us/

Basic Education Funding Commission. (2016). Factors used for the new formula. Retrieved from: http://basiceducationfundingcommission.pasenategop.com/files/2014/08/Basic-Education-Funding-Commission-Final-Report-Factors-Page-2015.pdf

Cantrell, S. & Kane, T. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study. Bill & Melinda Gates Foundation.* Retrieved from http://metproject.org/downloads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf

Collins, C. (2014). Houston, we have a problem: Teachers find no value in the SAS education value-added assessment system (EVAAS®). *Education Policy Analysis Archives, 22*(98). Retrieved from http://dx.doi.org/10.14507/epaa.v22.1594

Commonwealth of Pennsylvania. (2015). *Growth is not correlated with demographic variables*. Pennsylvania Department of Education. Retrieved from http://www.portal.state.pa.us/portal/server.pt/community/state_assessment_system/20965/p/1426518

Creswell, J.W. (2014). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Upper Saddle River, New Jersey: Pearson Education Inc.

Danielson, C. (2013). *The framework*. The Danielson Group. Retrieved from https://danielsongroup.org/framework/

Duncan, A. (2013). *Law and Guidance/ Elementary and Secondary Education*. United States Department of Education. Retrieved from http://www2.ed.gov/policy/eseaflex/secretary-letters/pa.html

EVAAS. (2015). Misconceptions about PVAAS reporting for Pennsylvania. Retrieved from: https://pvaas.sas.com/support/MisconceptionsAboutPVAASReporting.pdf

Gill, B., English, B., Furgeson, J., & McCullough, M. (2014). *Alternative student growth measures for teacher evaluation: Profiles of early-adopting districts*. (REL 2014–016). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory MidAtlantic. Retrieved from http://ies.ed.gov/ncee/edlabs

Goldhaber, D. & Theobald, R. (2013). *Do different value added models tell us the same things?* (Rev. ed.) Carnegie Foundation for the Advancement of Teaching. Retrieved from http://www.carnegieknowledgenetwork.org

Halle, T., Hair, E., Wandner, L., McNamara, M., & Chien, N. (2012). Predictors and outcomes of early vs. later English language proficiency among English language learners. *Early Child Res Q. 27*(1), 1-20

Harris, Douglas. (2012). *How do value-added indicators compare to other measures of teacher effectiveness?* Carnegie Foundation for the Advancement of Teaching.

Huang, F. L., & Moon, T. R. (2009). Is experience the best teacher? A multilevel analysis of teacher characteristics and student achievement in low performing schools. *Educational Assessment, Evaluation and Accountability, 21*(3), 209-234. doi:http://dx.doi.org/10.1007/s11092-009-9074-2

Huitt, W. (1996). Measurement and evaluation: Criterion- versus norm-referenced testing. *Educational Psychology Interactive*. Valdosta, GA: Valdosta State University. Retrieved from: http://www.edpsycinteractive.org/topics/measeval/crnmref.html

Ingvarson, L., Meiers, M. & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes & efficacy. *Education Policy Analysis Archives, 13*(10). Retrieved from: http://epaa.asu.edu/epaa/v13n10/

Isabel, L. A. (2010). *The effects of high quality professional development activities for teachers on students' tennessee comprehensive assessment program (TCAP) scores* (Doctoral dissertation). Retrieved from http://search.proquest.com (3404574)

Isenberg, E., Max. J, Gleason, P., Potamites, L., Santillano, R., Hock, H., & Hansen, M. (2013). Access to effective teaching for disadvantaged students: Executive summary (NCEE 2013-4002). Washington, DC: U.S. Department of Education.

Jackson, C.K., Johnson, R., Persico, C. (2014). *The effect of school finance reforms on the distribution of spending, academic achievement, and adult outcomes*. The National Bureau of Economic Research. Retrieved from http://www.nber.org.ezproxy2.library.drexel.edu/papers/w20118.pdf

James, L., Pate, J., Leech, D., Martin, E., Brockmeier, L., & Dees, E. (2011). Resource allocation patterns and student achievement. *International Journal of Educational Leadership Preparation*, *6*(4), 10. Retrieved from http://search.proquest.com

Koedel, C., & Betts, J. (2009). Value-added to what? How a ceiling in the testing instrument influences value-added estimation. National Bureau of Economic Research. Retrieved from: http://www.nber.org/papers/w14778.pdf

Kotte, D., Lietz, P., & Lopez, M. M. (2005). Factors influencing reading achievement in Germany and Spain: Evidence from PISA 2000. *International Education Journal, 6*(1), p. 113-124. Retrieved from http://eric.ed.gov.ezproxy2.library.drexel.edu/?id=EJ854960

McCaffrey, D.F., Lockwood, J.R., Koretz, D., Louis, T.A., Hamilton, L. (2004) Models for value-added modeling of teacher effects. *Journal of educational and Behavioral Statistics*, 2*9*(1), 67-101

Merz, E., He, X., Sowell, E.R., Noble, K.G. (2016). Age related differences in cortical thickness vary by socioeconomic status. Retrieved from: http://dx.doi.org/10.1371/journal.pone.0162511

Minitab. (2015). Put your data to work. Retrieved from http://www.minitab.com/en-us/

Morgan, G. B., Hodge, K. J., Trepinksi, T. M., & Anderson, L. W. (2014). The stability of teacher performance and effectiveness: Implications for policies concerning teacher evaluation. *Education Policy Analysis Archives*, *22*(95). Retrieved from http://dx.doi.org/10.14507/epaa.v22n95.2014

New America Foundation. (2014). *Federal education budget project*. Retrieved from http://febp.newamerica.net/background-analysis/no-child-left-behind-overview

Newton, X., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-Added Modeling of Teacher Effectiveness: An exploration of stability across models and contexts. *Educational Policy Analysis Archives*, *18*(23). Retrieved from: http://epaa.asu.edu/ojs/article/view/810

PA Bulletin. (2013). *Rules and regulations: Title 22 – education*. Department of Education. Retrieved from: http://www.pabulletin.com/secure/data/vol43/43-25/1115.html

Paige, R. (2002). *No child left behind: A desktop reference*. United States Department of Education. Retrieved from: https://www2.ed.gov/admins/lead/account/nclbreference/reference.pdf

Pennsylvania Association of School Business Officials. (2015). Special education funding. Retrieved from: http://www.pasbo.org/specialeducation

Pennsylvania Department of Education. (Rev.ed 2007). Pennsylvania's statewide system for school support. Retrieved from http://www.portal.state.pa.us/portal/server.pt?open=18&objID=353458&mode=2.

Pennsylvania Department of Education. (2012). Academic Achievement Report. About AYP. Retrieved from http://paayp.emetric.net/Home/About

Pennsylvania Department of Education (2013). ESEA flexibility request from Pennsylvania. Retrieved from: https://www2.ed.gov/policy/elsec/guid/esea-flexibility/map/pa.html

Pennsylvania Department of Education. (2015). Educator effectiveness systems in Pennsylvania. *Educator Effectiveness.* Retrieved from http://www.portal.state.pa.us/portal/server.pt/community/educator_effectiveness_project/20903

Pennsylvania Department of Education. (2016). Data and statistics. Retrieved from: http://www.education.pa.gov/Data-and-Statistics/Pages/default.aspx#tab-1

Pennsylvania Department of Education. (2016). Pennsylvania system of school assessment (PSSA). Retrieved from: http://www.education.pa.gov/K-12/Assessment%20and%20Accountability/PSSA/Pages/default.aspx#tab-1

Pennsylvania Public School Code. (2012). Public school code of 1949-omnibus amendments: Act of June 30, 2012, P.L. 684, no. 82. Retrieved from: http://www.legis.state.pa.us/WU01/LI/LI/US/HTM/2012/0/0082..HTM

Phillips, M.L. (2002). Second language learning. Retrieved from:

https://faculty.washington.edu/chudler/second.html

PVAAS Statewide Team for PDE. (2014). *PVAAS methodologies: Measuring growth and projecting performance.* Pennsylvania Department of Education. Retrieved from https://pvaas.sas.com/

PVAAS Statewide Team for PDE. (2015). Pennsylvania value added system: Guide to PVAAS public reporting. Pennsylvania Department of Education. Retrieved from: http://www.education.pa.gov/Documents/K-

12/Assessment%20and%20Accountability/PVAAS/Guides%20to%20Reporting/General%20Guides%20to%20Reporting/Guide%20to%20PVAAS%20Public%20Reporting.pdf

Raudenbush, S.W. (2015). Value added: A case study in the mismatch between education research and policy. *Educational Researcher 44*(2), 138–141. doi: 10.3102/0013189X15575345

Research for Action. (2013). *Teacher effectiveness: an update on Pennsylvania's teacher evaluation system*. Pennsylvania Clearinghouse for Education Research. Retrieved from http://www.researchforaction.org/wp-content/uploads/2013/12/RFA-PACER-Brief-on-Teacher-Eval-Dec-2013.pdf

Ready, D. D., & Lee, V. E. (2007). *Optimal context size in elementary schools: Disentangling the effects of class size and school size*. Brookings Papers on Education Policy, 99-135. Retrieved from: http://search.proquest.com/docview/757173602?accountid=10559

Rivers, J.C., Sanders, W.L., White, J.T., & White, P.S. (2010). SAS®EVAAS® statistical models. SAS Institute Inc. Retrieved from: http://www.sas.com/resources/asset/SAS-EVAAS-Statistical-Models.pdf

Sanders, W.L. & Horn, S.P. (1994). The Tennessee value-added assessment system (TVAAS): Mixed model methodology in educational assessment. *Journal of Personnel Evaluation in Education 8,* 299-311

SAS Institute. (2016). SAS EVAAS K-12 Statistical Models. Retrieved from: https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/sas-evaas-k12-statistical-models-107411.pdf

SAS Institute Inc. (2016). PVAAS Pennsylvania. Retrieved from: https://pvaas.sas.com/

SAS Institute. (2014). SAS® EVAAS® for K-12: Reliably assess and predict student performance. Retrieved from: https://www.sas.com/content/dam/SAS/en_us/doc/factsheet/education-evaas-104850.pdf

Sass, E. (2016). American educational history: A hypertext timeline. Retrieved from: http://www.eds-resources.com/educationhistorytimeline.html

Semrud-Clikeman, M. (2016). *Research in brain functioning and learning*. American Psychological Association. Retrieved from: http://www.apa.org/education/k12/brain-function.aspx

Survey Monkey. (2016). Make better decisions with the world's #1 survey platform. Retrieved from http://www.surveymonkey.com

Seton Testing Services. (2016). *The California achievement tests*. Retrieved from: http://www.setontesting.com/cat/

Springer, M. G. & Ballou, D. (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems. *Educational Researcher, 44*(2), 77–86. doi: 10.3102/0013189X15574904

Stout, W.L. (2013). *School level factors associated with Pennsylvania high schools that exceed PVAAS growth projections in reading* (Doctoral dissertation). Retrieved from: http://search.proquest.com. (3576497)

Stuit, D., Berends, M., Austin, M. J., & Gerdeman, R. D. (2014). *Comparing estimates of teacher value- added based on criterion- and norm-referenced tests* (REL 2014–004). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. Retrieved from http://ies.ed.gov/ncee/edlabs

UNICEF. (n.d.). *Early childhood development: The key to a full and productive life*. Retrieved from http://www.unicef.org/dprk/ecd.pdf

University of Pennsylvania. (2004). *Value-added assessment*. Center for Greater Philadelphia. Retrieved from http://www.cgp.upenn.edu/ope_value.html#8

U.S. Department of Education. (2009). *Race to the top program executive summary*. Retrieved from: https://www2.ed.gov/programs/racetothetop/executive-summary.pdf

U.S. Department of Education. (2012). *Flexibility to improve academic achievement and increase the quality of instruction. ESEA Flexibility.* Retrieved from http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html

U.S. Department of Education. (2012). *ESEA flexibility*. Retrieved from: http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html

U.S. Department of Education. (2011). *Race to the top awards*. Retrieved from: http://www2.ed.gov/programs/racetothetop/phase3-awards/pennsylvania.pdf

U.S. Department of Education. (2011). *Race to the top application for phase 3 funding*. Retrieved from: https://www2.ed.gov/programs/racetothetop/phase3-applications/pennsylvania-2.pdf

U.S. Department of Education (2015). Programs. Retrieved from: http://www2.ed.gov/programs/titleiparta/index.html

U.S. Department of Education (2016). Programs. Retrieved from: http://www2.ed.gov/programs/osepgts/index.html?exp=0

U.S. Department of Education. (2016). *Race to the top: Phase 3 grantees*. Retrieved from: http://www2.ed.gov/programs/racetothetop/phase-3-year-3-report.pdf

Vanderhaar, J. E., Munoz, M. A., & Rodosky, R. J. (2006). Leadership as accountability for learning: The effects of school poverty, teacher experience, previous achievement, and principal preparation programs on student achievement. *Journal of Personnel Evaluation in Education, 19*(1), 17. Retrieved from: http://search.proquest.com

Wahlstrom, D., Collins, P., White, T., & Luciana, M. (2010). Developmental changes in dopamine neurotransmission in adolescence: Behavioral implications and issues in assessment. *Brain and Cognition, 72*(1), 146-159. Retrieved from http://search.proquest.com

Wiley, E. W. (2006). A practitioner's guide to value added assessment: Educational policy studies laboratory research monograph. Retrieved from http://nepc.colorado.edu/publication/a-practitioners-guide-value-added-assessment-educational-policy-studies-laboratory-resea

Yoon, K.S., Duncan, T., Lee, S., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement*. (Issues & Answers Report, REL 2007-No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and regional Assistance, regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs

Zwerling, Howard. (2012). Comments on "Value-added estimates for phase 1 of the Pennsylvania teacher and principal evaluation pilot final report". Pennsylvania State Education Association. Retrieved from: https://www.psea.org/uploadedFiles/LegislationAndPolitics/Key_Issues/PSEACritique-MathematicaReport-March2012.pdf

Appendix A

Classroom Teaching Rating Tool Form

Commonwealth of Pennsylvania	DEPARTMENT OF EDUCATION	333 Market St., Harrisburg, PA 17126-0333

CLASSROOM TEACHER RATING TOOL FORM

PDE 82-1 (4/13)

Last Name	**First**	**Middle**
District/LEA	**School**	
Rating Date:	**Evaluation: (Check one)** ☐ **Semi-annual**	☐ **Annual**

(A) Teacher Observation and Practice

Domain	Title	*Rating* (A)	Factor (B)	Earned Points (A x B)	Max Points
I.	Planning & Preparation		20%		0.60
II.	Classroom Environment		30%		0.90
III.	Instruction		30%		0.90
IV.	Professional Responsibilities		20%		0.60
	(1)Teacher Observation & Practice Rating				3.00

Domain Rating Assignment 0 to 3 Point Scale (A)	
Rating	**Value**
Failing	0
Needs Improvement	1
Proficient	2
Distinguished	3

(B) Student Performance - Building Level Data, Teacher Specific Data, and Elective Data

Building Level Score (0 – 107)	
(2) Building Level Score Converted to 3 Point Rating	

(3) Teacher Specific Rating	
(4) Elective Rating	

(C) Final Teacher Effectiveness Rating – All Measures

Measure	Rating (C)	Factor (D)	Earned Points (C x D)	Max Points
(1) Teacher Observation & Practice Rating		50%		1.50
(2) Building Level Rating		15%		0.45
(3) Teacher Specific Rating		15%		0.45
(4) Elective Rating		20%		0.60
Total Earned Points				3.00

Conversion to Performance Rating	
Total Earned Points	**Rating**
0.00-0.49	Failing
0.50-1.49	Needs Improvement
1.50-2.49	Proficient
2.50-3.00	Distinguished
Performance Rating	

☐ **Rating: Professional Employee,** ***OR*** ☐ **Rating: Temporary Professional Employee**

I certify that the above-named employee for the period beginning ____________ (month/day/year) and ending ____________ (month/day/year) has received a performance rating of:

☐ **DISTINGUISHED** ☐ **PROFICIENT** ☐ **NEEDS IMPROVEMENT** ☐ **FAILING**

resulting in a FINAL rating of:

☐ **SATISFACTORY** ☐ **UNSATISFACTORY**

A performance rating of Distinguished, Proficient or Needs Improvement shall be considered satisfactory, except that the second Needs Improvement rating issued by the same employer within 10 years of the first final rating of Needs Improvement where the employee is in the same certification shall be considered unsatisfactory. A rating of Failing shall be considered unsatisfactory.

Date ______ Designated Rater / Position: ______ Date ______ Chief School Administrator ______

I acknowledge that I have read the report and that I have been given an opportunity to discuss it with the rater.
My signature does not necessarily mean that I agree with the performance evaluation.

Date ______ Signature of Employee ______

Appendix B

Letter to the Superintendent of School District

Naomi Bowen
701 Beversrede Trail
Kennett Square, PA 19348
610-389-0687
October 27, 2016

Dear Superintendent,

I am a doctoral student at Drexel University and writing to request permission to conduct research in your school district. The purpose of the research is to develop a greater understanding of how the use of the PVAAS Average Growth Index (AGI) in the Classroom Rating Tool impacts teacher perceived changes in the school environment. Additionally, the research will seek to understand the extent to which teachers understand how their PVAAS AGI is calculated and the extent to which they trust the validity of the calculation. The research will involve surveying all of the teachers of mathematics and reading at Middle School East and Middle School West.

Participation in the research study will be purely voluntary on the part of the teachers. Teachers will be informed of the purpose of the research and their rights in participating or declining participation. The survey will then be provided in electronically using Google forms..

Prior to conducting the research, the survey will need to be piloted among math and reading teachers in grades three through five for the purpose of determining if any adjustments need to be made to the survey prior to beginning the research. The pilot survey will also be provided electronically via Google forms.

The letter of consent for teachers and survey are attached. Thank you for taking the time to review these materials. I look forward to conducting research in your school district.

Sincerely,

Naomi Bowen

Appendix C

Pilot Survey Questions

Part 1: Demographic Information

Please circle the answer that best describes you.

1. Years of Teaching Experience: *0-5* *6-10* *11-15* *15+*

2. Content area for which you are a teacher of record: *Mathematics* *ELA*

Circle all that apply:

3. I have experience teaching students who are included in the following subgroups for my school:

   Special Education *Economically Disadvantaged* *Gifted*

4. I have discussed the validity of the PVAAS statistical methodology for measuring teacher effectiveness with colleagues.

   True *False*

Part Two: PVAAS Questionnaire

To what extent do you agree or disagree with the following statements?

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

1. I believe there is a relationship between teacher morale and receiving a PVAAS Average Growth Index on the Classroom Rating Tool. *Explain:*	1	2	3	4	5
2. I believe that teacher morale has become significantly more positive as a result of teachers receiving a PVAAS Average Growth Index on the Classroom Rating Tool. *Explain:*	1	2	3	4	5
3. I believe that teacher morale has become significantly more negative as a result of teachers receiving a PVAAS Average Growth Index on the Classroom Rating Tool. *Explain:*	1	2	3	4	5
4. I believe that receiving a PVAAS Average Growth Index in my Classroom Rating Tool has resulted in a change to the instructional practices I am expected to use in my classroom. *Explain:*	1	2	3	4	5
5. Receiving a PVAAS Average Growth Index in my Classroom Rating Tool has resulted in positive changes to the instructional practices that I am expected to use	1	2	3	4	5

in my classroom.

Explain:

6.	Receiving a PVAAS Average Growth Index in my Classroom Rating Tool has resulted in negative changes to the instructional practices that I am expected to use in my classroom.	1	2	3	4	5
	Explain:					
7.	Staffing changes have occurred as a result of a PVAAS Average Growth Index being calculated for my school building.	1	2	3	4	5
	Explain:					
8.	Scheduling changes have occurred as a result of a PVAAS Average Growth Index being calculated my school building.	1	2	3	4	5
	Explain:					
9.	The amount of support all students receive in mathematics and reading has increased as a result of a PVAAS Average Growth Index being calculated for my building.	1	2	3	4	5
	Explain:					
10.	The amount of support all students receive in mathematics and reading has decreased as a result of	1	2	3	4	5

the PVAAS Average Growth Index being calculated for my building.

Explain:

11. The amount of support struggling students receive in mathematics and reading has increased as a result of a PVAAS Average Growth Index being calculated for my building.	1	2	3	4	5

Explain:

12. The amount of support struggling students receive in mathematics and reading has decreased as a result of a PVAAS Average Growth Index being calculated for my building.	1	2	3	4	5

Explain:

13. I believe that the PVAAS Average Growth Index used in my Classroom Rating Tool is a valid measure of my effectiveness as a teacher.	1	2	3	4	5

Explain:

14. I believe that the PVAAS Average Growth Index used in my Classroom Rating Tool is influenced by factors outside of my control.	1	2	3	4	5

Explain:

15. I understand how my PVAAS Average Growth Index is calculated.	1	2	3	4	5

16. I understand the statistical algorithm utilized to calculate my PVAAS Average Growth Index score. 1 2 3 4 5

Appendix D

Survey

Consent Disclosure

Drexel University
Consent to Take Part In a Research Study

1. Title of research study: Measuring Teacher Effectiveness with the Pennsylvania Value-Added Assessment System

2. Researcher: Dr. Mary Jo Grdina and Naomi Rodriguez Bowen

3. Why you are being invited to take part in a research study

We invite you to take part in a research study because you are a teacher of mathematics or reading that receives an annual PVAAS AGI score for your content area.

4. What you should know about a research study

Someone will explain this research study to you.
Whether or not you take part is up to you.
You can choose not to take part.
You can agree to take part now and change your mind later.
If you decide to not be a part of this research no one will hold it against you.
Feel free to ask all the questions you want before you decide.

5. Who can you talk to about this research study?

If you have questions, concerns, or complaints, or think the research has hurt you, talk to the research team at: Contact Dr. Mary Jo Grdina, mfg29@drexel.edu, 215.895.2594 Naomi Rodriguez Bowen, na26@drexel.edu, 610-389-0687

This research has been reviewed and approved by an Institutional Review Board (IRB). An IRB reviews research projects so that steps are taken to protect the rights and welfare of humans subjects taking part in the research. You may talk to them at (215) 762-3944 or email HRPP@drexel.edu for any of the following:

Your questions, concerns, or complaints are not being answered by the research team.
You cannot reach the research team.
You want to talk to someone besides the research team.
You have questions about your rights as a research subject.
You want to get information or provide input about this research.

6. Why is this research being done?

The purpose of the research is to develop a greater understanding of how the use of the PVAAS Average Growth Index (AGI) in the Classroom Rating Tool impacts teacher perceived changes in the school environment. Additionally, the research will seek to understand the extent to which teachers understand how their PVAAS AGI is calculated and extent to which they trust the validity of the calculation.

7. How long will the research last?

We expect that you will be in this research study for 3 months.

8. How many people will be studied?

We expect about 30 people here will be in this research study out of 30 people in the entire study.

9. What happens if I say yes, I want to be in this research?

You will receive an electronic version of the survey to your school district email. The survey will take approximately fifteen minutes to complete.

10. What are my responsibilities if I take part in this research?

If you take part in this research, it is very important that you:

- Follow the investigator's or researcher's instructions.
- Tell the investigator or researcher right away if you have a complication or injury.

11. What happens if I do not want to be in this research?

You may decide not to take part in the research and it will not be held against you.

12. What happens if I say yes, but I change my mind later?

If you agree to take part in the research now, you can stop at any time it will not be held against you.

If you stop being in the research, already collected data may not be removed from the study database.

13. Is there any way being in this study could be bad for me?

There are no risks involved with participating in this study.

14. Do I have to pay for anything while I am on this study?

There is no cost to you for participating in this study.

15. Will being in this study help me in any way?

There are no benefits to you from your taking part in this research. We cannot promise any benefits to others from your taking part in this research.

17. What else do I need to know?

This research study is being done by Drexel University.

If you become ill during this study, please contact Naomi Bowen at telephone no. (610)-389-0687. If you require immediate medical attention, you should go to the nearest emergency room or call 9-1-1. It is important that you inform all emergency medical staff that you are participating in this study.

I have read the consent disclosure and agree to participate in this study ☐

Part 1: Demographic Information

Please circle the answer that best describes you.

1. Years of Teaching Experience: *0-5* *6-10* *11-15* *15+*
2. Content area for which you are a teacher of record: *Mathematics* *ELA*

Circle all that apply:

3. I have experience teaching students who are included in the following subgroups for my school:

   Special Education *Economically Disadvantaged* *Gifted*

4. I have discussed the validity of the PVAAS statistical methodology for measuring teacher effectiveness with colleagues.
   True *False*

Part Two: PVAAS Questionnaire

To what extent do you agree or disagree with the following statements?

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

1. I believe there is a relationship between teacher morale and receiving a PVAAS Average Growth Index on the Classroom Rating Tool.	1	2	3	4	5
Explain:					
2. I believe that teacher morale has become significantly more positive as a result of teachers receiving a PVAAS Average Growth Index on the Classroom Rating Tool.	1	2	3	4	5
Explain:					
3. I believe that teacher morale has become significantly more negative as a result of teachers receiving a PVAAS Average Growth Index on the Classroom Rating Tool.	1	2	3	4	5
Explain:					

4. I believe that receiving a PVAAS Average Growth Index in my Classroom Rating Tool has resulted in a change to the instructional practices I am expected to use in my classroom.	1	2	3	4	5
Explain:					
5. Receiving a PVAAS Average Growth Index in my Classroom Rating Tool has resulted in positive changes to the instructional practices that I am expected to use in my classroom.	1	2	3	4	5
Explain:					
6. Receiving a PVAAS Average Growth Index in my Classroom Rating Tool has resulted in negative changes to the instructional practices that I am expected to use in my classroom.	1	2	3	4	5
Explain:					
7. Staffing changes have occurred as a result of a PVAAS Average Growth Index being calculated for my school building.	1	2	3	4	5
Explain:					
8. Scheduling changes have occurred as a result of a PVAAS Average Growth Index being calculated my school building.	1	2	3	4	5
Explain:					
9. The amount of support all students receive in mathematics and reading has increased as a result of a PVAAS Average Growth Index being calculated for my building.	1	2	3	4	5
Explain:					
10. The amount of support all students receive in mathematics and reading has decreased as a result of the PVAAS Average Growth Index being calculated for my building.	1	2	3	4	5
Explain:					
11. The amount of support struggling students receive in mathematics and reading has increased as a result of a PVAAS Average Growth Index being calculated for my building.	1	2	3	4	5
Explain:					
12. The amount of support struggling students receive in mathematics and reading has decreased as a result of a PVAAS Average Growth Index being calculated for my building.	1	2	3	4	5
Explain:					
13. I believe that the PVAAS Average Growth Index used in my Classroom Rating Tool is a valid	1	2	3	4	5

measure of my effectiveness as a teacher.					
Explain:					
14. I believe that the PVAAS Average Growth Index used in my Classroom Rating Tool is influenced by factors outside of my control.	1	2	3	4	5
Explain:					
15. I understand how my PVAAS Average Growth Index is calculated.	1	2	3	4	5
16. I understand the statistical algorithm utilized to calculate my PVAAS Average Growth Index score.	1	2	3	4	5