2012

Accountability and Pennsylvania High Schools: Using a Value-Added Model to Identify, Quantify, and Track School Improvement

Todd Matthew Davies
Lehigh University

Accountability and Pennsylvania High Schools:

Using a Value-Added Model to

Identify, Quantify, and Track School Improvement

by

Todd Matthew Davies

Presented to the Graduate and Research Committee

of Lehigh University

In Candidacy for the Degree of

Doctor of Education

In the

Department of Educational Leadership

Lehigh University

May, 2012

Approved and Recommended for acceptance as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Education.

_____
Date

_____
Dissertation Advisor

_____
Accepted Date

Committee Members:

_____
Dr. Floyd Beachum, Chair

_____
Dr. Laurence Upton, Committee Member

_____
Dr. Bradley C. Landis, Committee Member

_____
Dr. George P. White, Committee Member

ACKNOWLEDGEMENTS

With gratitude, I acknowledge those who have supported, guided, mentored, and tolerated me while in pursuit of this degree, Doctor of Education.

Most notably and with the highest regard, I recognize my dear wife, Catherine. Her love, support, and patience have become my sustenance throughout this endeavor. Catherine and I joined this trial together and now, in writing these final words, I look forward to fully devoting myself to her and our *shared* goals, aspirations, and dreams.

I acknowledge my dissertation chair, Dr. Beachum, as well as my entire committee. I am immensely grateful that Dr. Beachum had the confidence in me to agree to chair this committee without having known me prior. Dr. White has been my primary advisor throughout my time at Lehigh. His guidance over the past five years has been highly influential with regard to both academic and professional interests. The Lehigh diploma, hanging in Dr. Landis's office, became my inspiration toward this ultimate goal. Since the beginning of this quest, and especially over the past year, his support, mentorship, and encouragement have become invaluable. Dr. Upton taught my first statistics class at Lehigh. Whether by coincidence or providence, four years later he became my chief source of counsel for this dissertation. I am deeply indebted to him for his persistence, long-suffering, and the (too) many surrendered evenings and weekends - relinquished on my behalf. Through this process he has become both mentor and friend. Per his request, I will pay this debt forward.

I acknowledge my family, friends, and co-workers who endured me through my seclusion, (occasional) irritability, and absence from meaningful personal and professional events. Finally, I would like to acknowledge my dad and mom who

imparted upon Kathy, Dave, Mark, and me, their children, the value of an education and the importance of faith. The completion of this academic journey and the composing of this dissertation are both reminders that *I can do all things through Him who strengthens me.*

DEDICATION

This dissertation and all that I am are dedicated to Catherine, my wife. Catherine, thank you for loving me, constantly praying for me, and your unconditional acceptance of me. I love you and thank God everyday for blessing me with such a wonderful wife, devoted friend, and dedicated ally in this world; you are my world.

TABLE OF CONTENTS

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

REFERENCES

APPENDICES

VITA

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This dissertation investigates the prevailing No Child Left Behind (NCLB) mandate as an effective platform to improve schools. The data compiled for use in this study represented 426 high schools in Pennsylvania and were retrieved from publicly accessible, state-sponsored sources. The statistical methodologies from the Pennsylvania Value-Added Assessment System (PVAAS) quantified the dependent measure as the 2010 Average Growth Index (AGI) in both Mathematics and Reading for each high school. The independent measure was each high school's School Improvement Classification (SIC). In this study, the SIC became a modification of each high school's NCLB school improvement status (ultimately determined by the number of consecutive years a school Made or Missed predefined Adequate Yearly Progress [AYP] targets).

A one-way ANOVA began the statistical analyses and detected significant differences among mean PVAAS scores across School Improvement Classification in both Mathematics and Reading. Post hoc analyses revealed that the 2010 Making Progress group outperformed other groups "Needing Improvement," thereby supporting the second research hypothesis in both content areas. However, the 2010 Making Progress group, as a whole, was unable to sustain its performance from 2010 to 2011, thus rejecting the final research hypothesis in both content areas.

The literature pertaining to the use of value-added methodologies (VAM) in public education generally describes the methodology used by PVAAS as an improvement over the existing, cross-sectional methodology to quantify school performance. Consistent with this viewpoint, the scores in Mathematics behaved in a somewhat predictable manner throughout all statistical tests. However, the Reading

PVAAS scores acted erratically, thus raising questions regarding the precision of VAM and the overarching legitimacy of its use to evaluate a school's effectiveness. From these findings, additional validation studies are recommended before relying solely on VAM to determine school performance.

The question remains as to whether the current system of external accountability in education is working, or whether the system is largely ineffective and quite possibly hampering improvement, especially in those schools where improvement is most needed. This manuscript concludes with recommendations for practice and further research.

CHAPTER 1

Introduction

No Child Left Behind (NCLB) imposes a prevailing system of external accountability over schools by measuring the achievement of students on standardized tests and issuing consequence to schools, based on the level of student performance (Tosh & Edwards, 2009). While many support the fundamental value of accountability in education (Davies, 2008; Raudenbush, 2004; Tosh & Edwards, 2009), there exists uncertainty regarding the effectiveness of legislated accountability to improve the quality of public schools. At the heart of this uncertainty are two unanswered questions regarding NCLB's approach to improving the quality of schools: (1) are schools accurately identified as needing improvement (Forte, 2010; Meyer, 2000; Reeves, 2005)? (2) Once identified, are schools supported at high enough levels to sufficiently increase student test scores and sustain improvement over time (Elmore, 2005; Forte, 2010)?

The first of these questions acknowledges the work of researchers that dispute the methodological soundness of NCLB as a school improvement platform (Forte, 2010; Linn, 2008; Wiliam, 2010). Essential to this first question is that a seemingly arbitrary percentage of students scoring proficient on high-stakes tests can effectively determine the overall or improved quality of a school (Linn, 2008; Meyer, 2000). Second, implicit to NCLB is the assumption that schools will instinctively know how and what to improve once identified. After all, NCLB's role in the school improvement process essentially ceases after identifying and consequentially sanctioning schools identified as failing. Therefore, the principals of failing schools are left to their own expertise to navigate the best path to increasing student achievement.

Marzano (2003a) asked and subsequently provided an answer to a popular question that permeates much of the school improvement literature, "What changes do we need to make in our schools and schooling, and how can we best implement those changes?" (p. 3). The literature regarding school improvement is replete with "best practices" of what schools can do to improve. For example, Marzano (2003b) produced a meta-analysis of over 35 years of research that identified the following school-level factors associated with effective schooling: a guaranteed and viable curriculum, challenging goals and effective feedback, parent and community involvement, safe and orderly environment, and collegiality and professionalism. Similarly, Hattie (2009) synthesized the research from over 800 meta-analyses and identified the following school-level factors associated with increased student learning: school size, small group learning, acceleration of curriculum, classroom management, classroom climate, and peer influences.

The literature on what schools can do to improve is vast and all encompassing. Yet, due to the influence of contextual variables when measuring school performance, much of the existing research is profoundly inadequate if its intended use is to guide principals through the process of leading school improvement (Raudenbush, 2004). Abelmann, Elmore, Even, Kenyon, & Marshall (1999) reported on the importance of school context when considering external accountability mandates in education:

> Nested within these developing external accountability systems are real schools: schools that have their own distinctive organizational characteristics and problems, schools that have unique student populations, schools situated in

diverse and particular communities and schools with their own institutional histories. (p.1)

Carpenter (2000) indicated that schools are elaborate systems, and reform based on the *silver bullets,* identified through much of the available research, is difficult to replicate due to the complex nature of each school. Schools are unquestionably complex. Even more complex however, is determining the quality or the improved quality of an individual school separate from that school's context. NCLB's cross-sectional "status" method for determining Adequate Yearly Progress (AYP) is found to be lacking, due to the confounding contextual variables at the school and student-level (Davies, 2008; Forte, 2010; Harris, 2010; Hershberg, Simon, & Lea-Kruger, 2004; Linn, 2008; Meyer, 2000; Raudenbush, 2004). For this reason, continued research in school improvement should begin by first unraveling the school effects from the contextual variables that correlate with student achievement.

Hershberg et al. (2004) reported that value-added assessment systems assist in separating the school effects from the contextual variables: "[Value-added assessment] is a far more accurate way to measure the performance of schools than absolute test scores. And though not a panacea, it gives educators an unprecedentedly rich opportunity to improve classroom instruction" (p.31). Harris (2010) reported, "Value-added measures are almost certainly better than our existing system of evaluating schools…" (p. 66). Value-added assessments take advantage of longitudinal data on individual students rather than looking at the cross-sectional performance of student cohorts to inform decisions about a school's improvement. Publicly available, value-added data are

currently used for AYP reporting purposes in several states including: Pennsylvania, Ohio, and Tennessee (Lockwood, McCaffrey, Stecher, Le, & Martinez, 2007).

Background of the Study

Since the launch of Sputnik in 1957, public schools have endured an ongoing stream of reform initiatives (Cuban, 2004). Precipitated by a threat to national security, the National Defense Education Act began the largesse of federal dollars spent on public education (Hunt, 2005). Conceptualized later as a "War on Poverty," President Johnson increased the role of the federal government in education during the succeeding decade by ratifying the Elementary and Secondary Education Act (ESEA) in 1965. The primary function of the ESEA was to increase educational opportunities for children in poverty by channeling federal dollars directly into schools serving high-poverty communities. In the 1970's, the influence of the federal government grew further with the passing of the Individuals with Disabilities Education Act (IDEA), known initially as the Education for All Handicapped Children Act (Tosh & Edwards, 2009). As the name implies, IDEA provided the guarantee of an appropriate public school education to all students and specifically addressed children with disabilities. The federal government galvanized its role in public schools in the 1980's, when The National Commission on Excellence in Education published, *A Nation at Risk*. Beyond its political significance as the first report of prominence from the newly authorized Department of Education (Hewitt, 2008), *A Nation at Risk* became the catalyst for a new, accountability-based reform agenda and paved the way for *Goals 2000: Educate America Act* and the reauthorization of ESEA.

"At the time of its signing, *Goals 2000* represented one of the greatest intrusions of the federal government into education policy" (Superfine, 2004, p. 10). *Goals 2000*

was the Clinton administration's response to the 1989 National Governor's Summit that initially provided both the urgency and framework for establishing national goals in education. The legislation's title provided a national deadline for the establishment of education goals in the following categories: (1) school readiness, (2) school completion, (3) student achievement and citizenship, (4) teacher education and professional development, (5) Mathematics and Science, (6) adult literacy, (7) safe and drug-free schools, and (8) parent participation (Superfine, 2005).

Where *Goals 2000* provided the framework for improvement, the reauthorization of ESEA in 1994, known as the Improving America's School Act (IASA), provided the monetary consequences to schools through changes to the Title I legislation (Superfine, 2004). Through IASA, the federal government began the enforcement of accountability by requiring all schools that received federal funds, through Title I, to employ certain aspects of the Systemic Reform Model (Superfine, 2005; Tosh & Edwards, 2009). Specifically, IASA required states to enforce the following accountability mandates for these schools: (1) development of state-wide common standards in Mathematics and Reading; (2) implementation of state-wide assessments to students in grades three, five, eight, and eleven; and (3) evaluation of school performance, determined by student scores on high-stakes tests (Forte, 2010). Even with these unprecedented and significant federal mandates, the IASA had three notable limitations that diminished its intended accountability over public schools. First, since each state has local control over education, IASA lacked the legitimate oversight to enforce the mandates at the school level. Second, since IASA targeted only those schools receiving federal dollars (i.e. Title I), its ability to influence student achievement was, at best, limited to those schools (Tosh

& Edwards, 2009). Third, IASA's accountability only addressed the school as the unit of improvement; it did little to decrease the widening achievement gap between minority, ethnic, and special needs subgroups. When President Bush reauthorized ESEA in 2001 as No Child Left Behind Act (NCLB), these limitations to true accountability were confronted through federal mandates at yet more unprecedented levels.

No Child Left Behind's Accountability

Of the many platforms espoused by this nation's elected officials, few engage the diverse American electorate more than education. The issues embodied by federal, state, and local policies toward schooling touch the lives of all Americans regardless of occupation, age, or political affiliation. The ubiquitous nature of education, coupled by the reported substandard performance of American students on international assessments (i.e., the Trends in International Mathematics and Science Study [TIMSS]; U.S. Department of Education, Institute of Education Sciences, 1999), created an ideal climate to reauthorize the ESEA with notable increases to the IASA's accountability mandates. The resulting No Child Left Behind Act of 2001 found bi-partisan support to strengthen accountability over public schools by addressing the learning outcomes of *all* children. The law was signed by President Bush in January 2002 and became actionable in July of the same year.

Abelmann et al. (1999) defined a working theory of accountability as the alignment of personal responsibility and shared expectations, in a system that has control over rewards and consequences. Newmann, King and Rigdon (1997) describe complete systems of accountability as having four discernible features: (1) established performance outcomes (i.e., test scores), (2) defined standards for success, (3)

consequences for performance such as sanctions or rewards to those identified as responsible, and (4) the authority to distribute rewards or sanctions. Nestled within NCLB's legislation, exist corresponding mandates aligned with a strong system of external accountability in public education. For example, NCLB requires states to: (1) adopt rigorous standards for Reading, Mathematics, and science; (2) administer (annual) high-stakes tests for Mathematics and Reading to all students in grades three through eight and at least one time between grades ten and twelve; (3) develop a definition for adequate yearly progress (AYP); and (4) employ a system to improve schools that do not meet annual targets. However, after nearly a decade of NCLB accountability, questions regarding its effectiveness at identifying and improving failing schools are yet to be answered.

Rigorous standards. Standards in education are academic or performance-based. Academic standards are the concepts and skills that ultimately define curriculum and become the criterion for high-stakes tests. Developing rigorous standards for instruction was the necessary first step in creating an external framework for accountability in education. A notable contribution of *A Nation at Risk* was establishing the initial urgency for creating academic standards (O'Shea, 2005). Consequentially, *Goals 2000* and IASA initiated the mandate for academic standards by requiring states to research, adopt, and employ Mathematics and language arts standards for one grade in elementary, middle, and high schools. NCLB expounded on IASA's groundwork with standards by adding more grade-levels and subject areas. At the time of this study, all states are required to (minimally) employ rigorous academic standards in grades three through eight and for at least one grade between ten and twelve (Tosh & Edwards, 2009). Even though NCLB

explicitly addressed the deficiencies of IASA, the manifestation of true accountability, including an improved international standing, is beset by the provision of each state's right to local control over education (see Reserved Powers Clause, U.S. Constitution, Article X). Internationally benchmarked content standards with corresponding common assessments (for all public schools) will ultimately be required to employ a fully functional system of accountability.

Performance standards refer to the level or "bar" at which a student, school, or district must perform in relation to high-stakes tests. The bar for students refers to the "cut score" needed to reach proficiency. At the time of this study, the bar for schools or districts is represented by the percent of students required to score proficient in order to reach its annual goal of AYP. Similar to content standards, each state is required to set performance standards for students, schools, and districts. However, unlike content standards, the federal government issued an unfathomable, one hundred percent proficiency ultimatum to all schools, districts, and state education agencies (SEA) by 2014 (Tosh & Edwards, 2009). Essentially, all schools must ensure that every student score proficient or higher on state tests or risk the likelihood of school and/or district-level sanctions. At the time of this study, certain states are receiving waivers from the 100% proficiency mandate, at the time of this study, no waivers have been issued for Pennsylvania.

Assessments. High-stakes testing is the hallmark of NCLB. Similar to the mandated grades and subjects for employing academic standards, students in grades three through eight and at least one time in high school are required to take standardized, high-stakes tests in language arts and Mathematics. Science is also a required test in

elementary, middle, and high school. However, the results from science assessments are not used to determine a school's performance for AYP.

Adequate yearly progress. The state-level accountability framework that connects NCLB's academic standards and assessments to student, school, and district performance standards is Adequate Yearly Progress (AYP). AYP is a complex monitoring system, designed to track each school and district's path to the ultimate goal of one hundred percent proficiency by 2014. In addition to proficiency on Mathematics and Reading assessments, AYP also measures student participation on high-stakes tests for both on Mathematics and Reading as well as the graduation rates for all schools with a graduating class (high schools). The complex nature of AYP increases markedly when the various metrics for calculating proficiency are used for both school outcomes as well as for all eligible student subgroups. Eligible subgroups include students from specific racial and ethnic backgrounds, students who are English Language Learners (ELL), students from Economically Disadvantaged (ED) homes, and Students with Disabilities (SWD).

School improvement. Schools that fail to meet the state's AYP requirements for two or more consecutive years are identified for "School Improvement." Consequences for schools identified for improvement are serious and increase each year a school fails to make AYP. Schools that continually fail to improve their AYP status are subject to the replacement of staff, leadership, or undergo a complete state take-over of the school or district. Table 1 represents the status and sanctions applied to schools that fail to make AYP in consecutive years. As proficiency standards for school and district performance continue to climb along the trajectory toward 100%, so does the likelihood that more, if

not all, schools will one day be placed in school improvement (McClure, 2005; Tosh & Edwards, 2009). With stakes this high, a concern exists as to why more directed efforts are not taken by the governing bodies to ensure NCLB is operating as it was intended - an accountability platform designed to improve public schools. Even more concerning is the absence of formal and accurate monitoring of the *improvement* of schools that are within the current system of accountability. After all, the true assessment of NCLB will ultimately be determined by the extent schools improve once identified as failing.

Pennsylvania System of School Assessment

The Pennsylvania System of School Assessment (PSSA) provides oversight for the accountability of Pennsylvania public schools. Due to high-stakes tests and accountability, the PSSA has become synonymous with the tests administered to Pennsylvania's students. Pennsylvania administers the Mathematics and Reading PSSAs in grades three through eight and eleven. The Science PSSA is used in grades four, eight, and eleven (beginning in 2007-2008). The Writing PSSA is employed in grades five, eight, and eleven. Even though high-stakes tests are the prominent determinant of AYP, they are not the only standard used to identify failing schools (Tosh & Edwards, 2009). NCLB requires all states to ensure that multiple measures are used to determine AYP. For Pennsylvania schools, the additional measures (besides student scores on high-stakes tests) include test *participation* rate and school attendance or high school graduation rate (if the school has a graduating class).

Table 1

Status and Sanctions for Years of School Improvement

Years failed to make AYP	Status/Sanction
Year 1	NCLB Status: Warning Develop two-year School Improvement Plan (SIP) by end of school year
Year 2	NCLB Status: School Improvement 1 Sanctions: • Employ two-year SIP • Technical assistance for the plan will be provided by the Local Education Agency (LEA) • Offer intra-district transfer option (Choice) for Title I schools only
Year 3	NCLB Status: School Improvement 2 Sanctions: • Continue o Employ/modify SIP; choice; technical assistance from LEA • Add o Supplemental education services (SES) -approved by State Education Agency (SEA)
Year 4	NCLB Status: Corrective Action – Year 1 Sanctions: • Continue o Employ/modify SIP; choice; technical assistance from LEA; SES • Add o Replace staff and or curriculum o Decrease autonomy o Outside expert assistance o Extend school year/days or restructure
Year 5	NCLB Status: Corrective Action – Year 2 Sanctions: • Continue o Employ/modify SIP; choice; technical assistance from LEA; SES • Add o Plan for restructuring (chartering, reconstitution, or privatization)

PSSA for high schools. Initial findings from research reported positive trends (improved test scores) for elementary schools since the inception of NCLB (Balfanz, Legters, West, & Weber, 2007; Duke & Jacobson, 2011). High schools, however, are unique and possess contextual barriers that make improvement more difficult to realize. Examples include the typically large size of high schools (compared to middle or elementary schools), the departmentalization of high school faculties, the age of high school students, and the accumulation of the negative effects from potentially ineffective elementary and middle school programs (Duke & Jacobson, 2011). An additional feature that is common to most Pennsylvania high schools is that the PSSAs are administered only once, in eleventh grade. Assessing students only one time in high school, rather than at the end of each year or at the end of a course, creates the need for high schools to employ local assessments, aligned to the PSSA, to measure the impact of instruction on student learning during non-tested years.

AYP measures student outcomes in high schools across three indicators, (1) performance and (2) participation on the eleventh grade Mathematics and Reading PSSAs, and (3) graduation rates (Pennsylvania Department of Education [PDE], 2010a). Pennsylvania high schools are held accountable for all students, including eligible subgroups of students. Eligible subgroups include students from specific racial and ethnic backgrounds, students who are English Language Learners (ELL), students from Economically Disadvantaged (ED) homes, and Students with Disabilities (SWD). For student subgroups to be large enough to determine AYP, they must contain 40 or more students ($n \geq 40$).

Rigorous standards. Pennsylvania high schools employ criterion-referenced assessments, aligned to rigorous state academic standards, to measure student achievement in tested subject areas. Students attain scaled scores in each assessed subject that are associated with one of four performance levels, below-basic, basic, proficient, and advanced. AYP is determined by calculating the percent of students scoring proficient or advanced on the Mathematics and Reading PSSA. Like all states, Pennsylvania established its own trajectory for the 2014 goal of 100% proficiency in Mathematics and Reading. Table 2 represents the annual proficiency targets for determining AYP in Mathematics and Reading.

Table 2

Pennsylvania Proficiency Targets

Year	2002-2004	2005-2007	2008-2010	2011	2012	2013	2014
Mathematics	35%	45%	56%	67%	78%	89%	100%
Reading	45%	54%	63%	72%	81%	91%	100%

Graduation rate. All Pennsylvania high schools with a graduating class are required to use graduation rate as a target for AYP. Beginning during the 2012 AYP reporting year (using the 2011 graduating seniors), Pennsylvania will fully employ the mandated 4-year Cohort Graduation Model to calculate the high school graduation rate. This new method for calculating high school graduation rates will determine its AYP status for all students as well for every eligible subgroup ($n \geq 40$). The target for high school graduation in Pennsylvania is 82.5%. For schools that do not meet the target, they would need to demonstrate a ten percent reduction of the difference between the previous year's graduation rate and 85% to make AYP.

Participation rate. At least 95% of students overall and within each measurable subgroup must take the test.

Adequate yearly progress. For Pennsylvania high schools to make AYP, all measures (achievement, participation, and graduation) must be met for all students in grade eleven as well as eligible subgroups of students ($n \geq 40$). Depending on the number of eligible subgroups, a high school may have as many as 40 indicators (for participation and performance) to determine AYP status. Missing a single indicator would prevent a school from making AYP for that year. Missing an indicator during the subsequent year would cause the school to become placed in school improvement. Schools will either meet the minimum proficiency thresholds for AYP in a given year or not. For schools that do not meet the minimum thresholds for proficiency, Pennsylvania received approval from the federal government for schools to make AYP through alternate means. For example, schools can make AYP through "Safe Harbor" by reducing the number of non-proficient students by 10%. In this sense, Safe Harbor represents a cross-sectional "growth" model as a means to make AYP.

Growth model and value-added. During the 2008-2009 school-year, Pennsylvania updated its accountability workbook to include an additional Growth Model for determining AYP status. In Pennsylvania, the Growth Model uses the Pennsylvania Value-Added Assessment System (PVAAS) to project the proficiency of a student on future PSSA assessments to determine AYP for the current year. For example, if a student scores below basic or basic during the current school year but is projected to score proficient in an upcoming year (see Table 3 for projections) then that student's score would be adjusted to proficient to calculate AYP during the current year. The

adjusted score would be applied to all subgroups associated with that student. However, Pennsylvania high schools cannot employ the Growth Model for AYP purposes since eleventh grade students are unable to project proficiency beyond the current year.

In addition to projecting an individual student's growth, PVAAS can also isolate the school's contribution to student learning through the estimated Growth Measure (GM). Essentially, PVAAS untangles the student-level context variables from the value-added estimates of the school to provide a numerical representation of the school's effect on a student's learning. For high schools, the GM is the difference between the mean eleventh grade students' score in Mathematics or Reading and their mean predicted scores.

The GM, for the most recent year, divided by its standard error determines the "Average Growth Index" (AGI) for each school. The standardized AGI is used to compare all Pennsylvania high schools. Though imperfect, the value-added methodologies employed by PVAAS are considered fairer than cross-sectional methods when measuring school performance or improvement on high-stakes tests (Hershberg, Simon, & Lea-Kruger, 2004; Harris, 2010; Linn, 2008; Meyer, 2000).

Table 3

Projected Grades for Growth Model

Tested Grade	Projected Grades for Proficiency
3	Actual Grade 3 Scores
4	Projected Scores for Grade 6
5	Projected Scores for Grade 7
6	Projected Scores for Grade 8
7	Projected Scores for Grade 8
8	Projected Scores for Grade 11
11	Actual Grade 11 Scores

School improvement. According to NCLB, schools that fail to make AYP for two or more consecutive years are identified for "School Improvement" and required to complete and submit a school improvement plan (SIP). Pennsylvania requires all failing schools to develop and submit a two-year improvement plan beginning the first year of not making AYP. The school's improvement plan, in addition to increases in sanctions and technical assistance, will continue each consecutive year the school does not make AYP.

Purpose

According to Michael Fullan, "accountability and improvement can be interwoven" (as cited in Schmoker, 1999, p. 53). Implicit with Fullan's statement is that the system of accountability employs an effective measure of performance to determine improvement. The purpose of this study is to make use of value-added methodologies to determine if the school improvement process, as defined by NCLB, is effectively improving student achievement in Pennsylvania high schools. Specifically, this study will determine whether a high school's PVAAS AGI differs between years and levels of school improvement for both Mathematics and Reading. Simply, this study will determine if the policy mandates of NCLB (i.e. sanctions, rewards, and school improvement plans) are effectively improving Pennsylvania's high schools. This study has implications for teachers, school and district leaders, state education agencies, university professors, and policy makers who require answers to the questions regarding NCLB's effectiveness at improving schools.

Research Questions

No Child Left Behind is a federally mandated accountability system dedicated to improving public schools. Accountability in education is not possible without accurately measuring the performance of schools. The current cross-sectional "status" model used by NCLB is unable to isolate the contributions made by schools due to confounding context variables and is therefore ineffective at measuring the improvement of schools. The ineffective methodology concerns the principals leading schools through a mandated school improvement process. Many consider that this issue has a heightened sense of urgency for high schools due to the increased time it would take to fully employ meaningful changes at that level (Duke & Jacobson, 2011). Therefore, this study will seek answers to the following research questions regarding NCLB's overall effectiveness at improving the performance of schools identified as failing:

Question 1a: Does the PVAAS Average Growth Index (AGI) in Mathematics differ between Pennsylvania high schools across School Improvement Classifications (SIC)?

Question 1b: Does the PVAAS Average Growth Index (AGI) in Reading differ between Pennsylvania high schools across School Improvement Classifications (SIC)?

Question 2a: Do high schools identified as Making Progress demonstrate higher PVAAS Average Growth Indices (AGI) in Mathematics than high schools Needing Improvement?

Question 2b: Do high schools identified as Making Progress demonstrate higher PVAAS Average Growth Indices (AGI) in Reading than high schools c?

Question 3a: Do the PVAAS Average Growth Index (AGI) gains in Mathematics for high schools identified as Making Progress continue to subsequent testing years?

Question 3b: Do the PVAAS Average Growth Index (AGI) gains in Reading for high schools identified as Making Progress continue to subsequent testing years?

Definition of Variables

2010 PVAAS Average Growth Index: The Average Growth Index (AGI) is the dependent variable. For high schools in this study, the AGI is the Growth Measure (GM) for the most recent year divided by its standard error. This is calculated for both Mathematics and Reading, separately (PVAAS Statewide Core Team for PDE, 2011).

School Improvement Classification (SIC): For this study, the independent variable is a modified version of the 2010 NCLB School Improvement status for high schools in Pennsylvania. Modifications to the independent variable are discussed in Chapter Three.

Definition of Terms

Cross-Sectional Status: Cross-sectional status refers to the current model employed by NCLB to determine AYP, where the improvement of a group or school is measured by comparing different cohorts of students (i.e., comparing the status of 2009 cohort of eleventh grade students to the status 2010 cohort of eleventh grade students).

External Accountability: External accountability refers to the policies, mandates, and mechanisms, beyond the school, that measure and enforce student performance outcomes against specific targets (Cobb & Rallis, 2008).

Growth Measure: Growth Measure is critical to computing the Average Growth Index (AGI), used as the dependent variable in this study. For eleventh grade Mathematics and Reading the Growth Measure uses a predictive methodology and is:

> …a function of the difference between the students' observed scores (Avg. PSSA Score) and their predicted scores (Average Predicted PSSA Score). If students score as expected (i.e., students' observed scores are equal to their predicted scores), the estimated District/School Effect would be 0, indicating progress similar to the average district/school in the state. In other words, the Growth Measure is the amount of progress made by that group of students. (PVAAS Statewide Team for PDE, 2011, pp. 23-24)

High Schools: High schools for this study are defined by the following: (1) had a graduating class during the 2010 school year, (2) had PVAAS AGI scores reported for both Mathematics and Reading reflecting the eleventh grade assessment during the 2010 assessment year, (3) does not contain grade eight or below, (4) is not a private school, (5) is not a charter school, (6) is not a school operated by an intermediate unit, (7) is not a vocational-technical school, (8) is not state-owned, (9) is not a special program school, (10) is not a juvenile detention school, (11) is not a private academy, (12) is not an approved private school, and (13) has at least two years of prior eleventh grade assessment data.

Pennsylvania Value Added Assessment System (PVAAS): PVAAS is Pennsylvania's statistical analysis system that uses longitudinal data of students' performances on the

PSSA Mathematics and Reading assessments to predict and determine the growth of schools. A school's growth can be compared to other schools using the Average Growth Index (AGI).

Significance of Study

The findings from this study will have important implications regarding the methodologies used to measure the improvement of failing high schools. At the time of this study, NCLB mandates cross-sectional status comparisons of the percent of students and student subgroups scoring proficient on high-stakes tests to make decisions regarding a school's performance. However, these types of comparisons fail to weigh the contributions of the school apart from the context variables associated with the student (i.e., student race, gender, family income), thus providing little evidence of a school's overall or improved performance (Hershberg et al., 2004; Topping & Sanders, 1999). The value-added methodologies employed by PVAAS give promise to a more accurate determination of school performance by controlling for contextual variables that are beyond the school's influence (Harris, 2010; Topping & Sanders, 1999). Accurately measuring the performance of a school is a requisite to identifying schools that are failing or improving. Moreover, accurately measuring the improved performance of individual schools, identified as failing, will provide evidence regarding the overall effectiveness of NCLB's accountability.

Limitations

This study is classified using Johnson's (2001) description of non-experimental, quantitative research. Johnson asserts a typology where all non-experimental, quantitative research is classified across two dimensions, time and research objective. The present study is considered *retrospective* with regard to the dimension of time and *explanatory* with regard to research objective. According to Fraenkel and Wallen (2009), researchers should exercise caution when reporting or inferring causal relationships with non-experimental studies due to the inherent lack of internal validity. Fraenkel and Wallen describe two threats to all non-experimental studies, (1) the lack of randomization and (2) inability to assert control over the independent variable. This study's large, nearly exhaustive, sample of participating Pennsylvania high schools and the use of existing data will serve to decrease the extent that these threats can obscure findings. Moreover, rather than testing a *new* hypothesis, this study's aim is to quantify the intended relationship between external accountability and school improvement using value-added methodologies.

The sample for this study includes all Pennsylvania public high schools. Therefore, the results from this study should not be generalized beyond the boundaries of Pennsylvania. However, since all states are influenced by NCLB and required to employ high-stakes tests, it is not unreasonable to surmise that the findings from this study would match the findings in other states with similar state accountability policies.

The participant high schools used in this study will ultimately be limited to the schools that have results reported by the Pennsylvania Department of Education (PDE) on the PVAAS public reporting website at: https://pvaas.sas.com/evaas/signin.jsf.

The Pennsylvania Department of Education (PDE) does not archive PVAAS AGI records from year to year. Therefore, the 2010 high school AGI scores used in this study were retrieved from the PVAAS public reporting website at https://pvaas.sas.com/evaas/signin.jsf, three months prior to performing the analysis. To access the historical assessment records, potential researchers must complete the agreement attached as Appendix A. It is important to note that this study only represents a snapshot of Pennsylvania's high schools, additional research using the data from future years is recommended to further substantiate the findings from this study

CHAPTER 2

Review of Literature

The purpose of this chapter is to provide a summary of the relevant research in the area of accountability and public education as well as validation of variables used in this study. This chapter is reported in two sections. The first section provides a narrative of key terms and relevant research pertaining to accountability in education. The second section addresses the theory-base and existing research regarding Value-Added Modeling (VAM), specifically targeting Pennsylvania's Value-Added Assessment System (PVAAS). A summary is provided at the end of each section.

Accountability in Education

Accountability in education has evolved over the years into a diverse concept leading researchers, politicians, practitioners, and the general-public to define it with significant professional and personal latitude. A query of the phrase, "accountability in education" returns a compendium of literature regarding educational theories, types, terms, mechanisms, and taxonomies. According to Michael Heim (1996), the meaning of accountability in education is "muddled," referring to "responsibility," "oversight," and/or "compliance" (p. 2). Henry Levin's (as cited in Kirst, 1990) appraisal of more than 4,000 pieces of accountability literature concluded that the "concept of accountability [in education] is vague and rhetorical" (p. 6). Levin specifically addressed the absence of a common definition and the requirement for a consistent framework needed to organize the vast array of techniques the phrase implies. Levin's research was conducted in 1972 (Kirst, 1990). Notwithstanding the equivocality of the phrase, the concept and consequence of accountability in education continues to pervade the

political, academic, and public discourse. To that end, a circumspect review of relevant literature will be provided with the aim of delimiting the concept of educational accountability to the context of school improvement.

Internal and External Accountability

Accountability operates either internally or externally. External accountability depicts an authority outside the school or district while internal accountability denotes an internal source (Elmore, 2005). Research describes a strong relationship between internal and external accountability systems, where schools with high levels of internal accountability have greater success responding to systems of external accountability (Abelmann et al., 1999; Elmore, 2005; Fuhrman, 1999; Newmann, King, & Rigdon, 1997). However, it is not enough for schools to simply attain high levels of internal accountability, the alignment between the two systems is key to the success of either system (Fuhrman, 1999). Abelmann et al. (1999) posits that internal accountability mechanisms mediate the effects of external accountability. Therefore, the relationship between legislated (external) accountability and school improvement is arguably the school's ability to develop a system of internal accountability and effectively align that system to external accountability mechanisms.

Theory of Accountability

Kirst (1990) defines accountability, literally, as a relationship between an authority and steward, where someone in authority has the power to remove a steward based on a standard of performance:

> At the heart of the process is the party "standing to account," the steward, to explain as rationally as possible the results of efforts to achieve the specified tasks or objectives of his stewardship. (p. 14)

The hierarchical relationship between the authority and steward is an established concept of accountability in education (Darling-Hammond, 2004; Heim, 1996; Kirst, 1990). For example, elected board members (stewards) are accountable to an electorate, laws and policies that govern education (authority); district officials (stewards) are accountable to board members, parents, laws and policies (authority); principals (stewards) are accountable to district officials, parents, laws and policies (authority). The Principal-Agent Theory is one framework that espouses certain aspects of this relationship.

Also referred to as Agency Theory, Principal-Agent Theory regards working relationships in diverse contexts as a series of contracts between members, where one member (the principal) delegates or "contracts out" a task to another member (agent) due to a lack of time or expertise on the part of the principal (Wohlstetter, Datnow, & Park, 2008). As a framework, Principal-Agent Theory has been applied to diverging contexts in multiple disciplines, including: management (employer-employee), political science (legislator-bureaucrat), and economics (buyer-supplier; Wohlstetter et al., 2008). Additionally, educational researchers applied the same theoretical framework to an assortment of studies including: data driven decision-making (Ferris, 1992; Wohlstetter, et al., 2008), goal setting (Davies, Coates, Hammersley-Fletcher, & Mangen, 2005), and political power structures in education (Vanhuysse & Sulitzeanu-Kenan, 2009). Notwithstanding the promise this theory brings to education as a highly adaptable framework, the voluntary nature of the contract between principal and agent and the

assumption of information asymmetry, favoring the agent, preclude its utility as an intact framework for this present study. However, the context of the principal-agent relationship is conceptualized in the current bureaucratic mechanisms employed through NCLB and therefore noted within this present study.

Certainly, other frameworks in education exist that address aspects of accountability in education. However, the present study addresses the relationship between legislated accountability and school improvement. Elmore (2005) describes most external accountability systems as "primitive" and lacking a theoretical base in school improvement when he stated, "schools are expected to improve their performance over time, as measured by external tests. Just how this occurs, what it entails, and the factors determining progress are not specified" (p.138). Abelmann et al. (1999) presented a working theory of accountability as a cross-section of personal responsibility, shared expectations, and internal and external accountability mechanisms by asserting, "the power of external accountability systems is a function of the alignment between the norms and values of these systems and the internal mechanisms of a school" (p. 6). Working from this theory-base, the present study will construct a conceptual framework that depicts the guiding mechanisms of accountability necessary to improve public high schools.

Mechanisms of Accountability

Many researchers refer to the mechanisms of accountability by the "type," "approach," "source," "authority," or "strategy" (Darling-Hammond, 1989; Darling-Hammond, 2004; Heim, 1996; Kirst, 1990; Newmann, King, & Rigdon; 1997; Wohlstetter, 1991). Two prominent studies provide taxonomies for the mechanisms of

accountability in education. Kirst (1990) adding to the research from Levin (as cited in Kirst, 1990) examined accountability through six mechanisms: performance reporting, monitoring and compliance with standards, incentive systems, reliance on market, changing the locus of control, and changing professional roles. Darling-Hammond (1989) interpreted accountability in education as a result from political, legal, bureaucratic, professional, and/or market-based mechanisms. Unlike Kirst (1990), who denotes accountability mechanisms more broadly, referring to both authority and strategy, Darling-Hammond (2004) clearly depicts the authority or the source of accountability when referring to mechanisms. The design and purpose of the present study aligns more closely with specific mechanisms outlined by Darling-Hammond (1989).

Mechanisms of accountability are fluid and originate from different theoretical frames and settings (Heim, 1996). Each mechanism has strengths and weaknesses, depending on the goal or intended purpose of the accountability. The following are descriptions of accountability mechanisms as well as implications for this current study.

Political accountability. Political accountability refers to the manner by which local, state, and federal representatives are held accountable in a democratic society (Heim, 1996). Political mechanisms of accountability in education began moving away from the state and local electorate to the national electorate with the initial passage of Elementary and Secondary Education Act (ESEA) in 1965. Strong, political mechanisms of accountability, even at the national level, are suitable and even beneficial for passing general policy directions (Darling-Hammond, 2004). However, a national voting constituency brings a greater heterogeneity of beliefs and expectations that may weigh on

the intended effects of accountability and expediency of goal attainment. This pluralism is evidenced by the discourse among societal groups that, on one hand, credit current accountability policies that require all schools to be held accountable for all subgroups of students while also criticizing the terms of the policy as they apply to equity across subgroups or the method of goal attainment (Hanushek & Raymond, 2005; Murnane, 2007).

Legal accountability. Legal mechanisms of accountability in education grant individuals the authority to hold various aspects of public education accountable through the due process of law. Local school board policies, Pennsylvania's School Code, and Individuals with Disabilities Education Act (IDEA) are all examples of legal mechanisms of accountability in public education. Legal accountability is highly valued in a free society and is effective as long as all members of society have access to the court system. Legal accountability is limited by the high costs of maintaining an effective monitoring system and often uses punishments to induce the compliance of members in the system (Heim, 1996).

Bureaucratic accountability. The promulgation of rules and regulations from various levels of federal, state, and local education agencies characterizes the complexity of bureaucratic accountability (Heim, 1996). By nature, this mechanism of accountability emphasizes a hierarchical, subordinate-superior association where subordinate positions are accountable to superiors that hold expectations of them and power over them (Cobb & Rallis, 2008). This mechanism for accountability is most recognizable within the current system of public education where bureaucratic structures decentralize the decision-making and accountability across a hierarchy. The complexity of bureaucratic

accountability postulates the need for accountability to be exercised across multiple relationships. For example, as NCLB holds the school as the "unit" for improvement, principals of schools are accountable for both outcomes of students (test scores) and the inputs from teachers (evaluations of instructional delivery). Moreover, building principals are accountable to multiple stakeholders, including: students, parents, teachers, district administrators, state requirements, and federal mandates. The numerous rules and regulations needed to maintain a fully functional bureaucratic system of accountability typically reduce the autonomy of each school by forcing compliance to general policy mandates (Heim, 1996).

Market-based accountability. Market-based accountability uses a free-market system, predicated on a range of choices, to encourage quality improvements among public schools (Heim, 1996). Market-based accountability continues to emerge as a potential mechanism for improving public schools. The guiding principal behind market-based mechanisms is that tax dollars follow the student to the school of their choosing, which makes losing that income a significant consequence to the school losing the student (Manno, 2004). Two present-day examples of market-based mechanisms in public education include competitive charter (including cyber-charter) schools and voucher programs. NCLB also promotes a market-based concept through the school improvement process known as "School Choice." School Choice is a mandated sanction that permits students from a failing school to attend another school in the district at the district's expense.

Professional accountability. Professional accountability uses the values promoted by a group of professionals to establish the standard of accountability (Heim,

1996). Darling-Hammond (1989) suggested this mechanism of educational accountability based on the teachers' competence and effectiveness rather than organizational rules and accompanying procedures. Essentially, professionalism and accountability have shared principles, "Professional prerogatives to make decisions are accompanied by professional obligations to do so in a responsible manner" (p. 15). Internationally, professional accountability mechanisms are employed with varying accounts of effectiveness in Japan (Darling-Hammond, 1989; Wong, 2003) and Germany (Fried, 2009). Professional accountability gained the attention of district leaders and researchers who view Professional Learning Communities (PLC) as a means of holding professionals accountable to the goals and values represented by the group of participating professionals (Stoll, Bolam, McMahon, Wallace, & Thomas, 2006).

Research on Accountability in Education

Forte (2010) explained the logic behind NCLB's accountability as the following: (1) Define what students should know and be able to do, (2) assess students using tests aligned to the same standards, (3) use the scores from the assessments to determine the school's effectiveness and make decisions meant to improve school functioning. Wiliam (2010) defined the logic of current accountability practices as simply: "students attending higher quality schools will (by definition) have higher achievement than those attending lower quality schools" (p. 110). Hochberg and Desimone (2010) explained that the present national accountability policy (using high-stakes tests) is designed to pressure schools into making meaningful changes that lead to improvement. After nearly ten years of implementing NCLB policies, the scholarship regarding the effects of state and federal legislation on student achievement yields varying results and often cultivates

more questions than answers (Lee, 2006). Small sample studies and case studies saturate the existing literature (Chatterji, 2002; Harris & Herrington, 2006; Lee & Wong, 2004). Rigorous studies within individual states are less common, yet necessary to provide valuable information to researchers regarding accountability policies specific to each state.

Educational researchers have long debated how accountability policies at the state level influence student outcomes (Carnoy & Loeb, 2002; Hochberg & Desimone, 2010; Lee, 2006; Lee & Wong, 2004). Several researchers found positive effects attributable to state accountability policy on student outcomes (Carnoy & Loeb, 2002; Hanushek & Raymond, 2005; Lee & Wong, 2004). For example, Carnoy and Loeb (2002) found positive and significant relationships between the strength of states' accountability systems and gains on the National Association of Educational Progress (NAEP) eighth-grade mathematic assessments (Carnoy & Loeb, 2002). In a similar study, Hanushek and Raymond (2005) reported a significantly positive relationship between state accountability policies and student gains on the NAEP mathematic and Reading assessments. However, unlike the research conducted by Carnoy and Loeb (2002), who found significant discoveries based on the strength (consistency) of the state's accountability policies, Hanushek and Raymond (2005) reported significant findings for the state policies that attached consequences to student outcomes. As the issuance of consequences is a component of accountability, several researchers studied the interaction of sanctions and student achievement, directly.

Amrein and Berlinger (2002) sought to determine the transferability of performance between high-stakes assessments and other assessments in states with strong

accountability systems (as determined by sanctioning of schools for student outcomes). The framework for their research is illustrated by two assumptions. First, "if the high-stakes testing of students really induces teachers to upgrade curricula and instruction or leads students to study harder or better, then scores should also increase on other independent assessments" (p. 22). Secondly, Amrein and Berlinger (2002) assume that since sanctioning is a component of accountability, then faster growth (i.e., increases in student assessment scores) would be observed after high-stakes accountability policies were introduced.

Amrein and Berlinger (2002) compared the assessment scores from a sample of 18 states that administered "the most severe" (p. 19) consequences for high stakes testing to the national average to determine whether sanctioning *schools* leads to increases in *student* performance on other, low-stakes assessments. In this sense, Amrein and Berlinger (2002) considered each state's policy toward administering consequences relative to student performance on high-stakes tests as the independent variable and the scores from the following four national assessments as dependent measures: NAEP, American College Testing programs (ACT), Scholastic Achievement Test (SAT) and Advanced Placement (AP) assessments. Table 4 represents the specific assessments, grade levels, and years used for the study.

Through the study design, Amrein and Berliner (2002) plotted each state's scores from each assessment and grade and compared the net gain (+ or -) to the national average. For example, Mississippi was provided with the baseline year of 1988 (the first year Mississippi mandated graduation exams) and gained a total of 8 points from 1992-2000 on the eighth grade NAEP Mathematic assessment (no data were available prior to

1992). However, since the national average also gained 8 points during the same period of time, the gain was considered "neutral" by the researchers. All of the sampled states were plotted in a similar fashion using the four assessments for the tested grades.

Table 4

Data Sources for Amrein & Berliner

Assessment	Data source	Years
ACT	Composite scores	1980-2001
ACT	Participation rates	1994-2001
SAT	Composite scores	1977-2001
SAT	Participation rates	1991-2001
AP	Grade 11/12 who took AP tests	1991-2000
AP	Grade 11/12 earning a (3) or higher on AP	1995-2000
NAEP	Grade 4 Mathematic composite scores	1992, 1996, 2000
NAEP	Grade 8 Mathematic composite scores	1990, 1992, 1996
NAEP	Grade 4 Reading composite scores	1992, 1994, 1998
NAEP	Grade 8 Reading composite scores	1998

Note. Adapted from "High-Stakes Testing, Uncertainty, and Student Learning" by A. L. Amrein & D.C. Berliner, 2002, *Education Policy Analysis Archives, 10*(18), p.p. 27-28.

Amrein and Berlinger (2002) concluded that even though individual states reported increases on high-stakes tests, the transfer of that learning to other measures in comparable domains remains unchanged.

> At the present time, there is no compelling evidence from a set of states with high-stakes testing policies that those policies result in transfer to the broader domains of knowledge and skill for which high-stakes test scores must be indicators. Because of this, the high-stakes tests being used today do not, as a

> general rule, appear valid as indicators of genuine learning, of the types of learning that approach the American ideal of what an educated person knows and can do. (Amrein & Berliner, 2002, p. 64)

Amrein and Berlinger (2002) also concluded that scores on national assessments in many cases decreased as a result of consequences imposed by accountability policies at the state level (Amrein & Berlinger, 2002; Rosenshine, 2003).

The findings from Amrein and Berlinger (2002) are problematic based on methodological principles. For example, the researchers identified the national composite of all states as a comparison group rather than creating a group of states not employing high-stakes tests as the study's control group. This issue is of concern since almost one-third of the national, comparison-group sample is comprised of states with strong high-stakes accountability policies. Therefore, an argument could be made that the aggregation of high-stakes testing policies, within the national sample, could have contributed to the overall increases in scores.

Rosenshine (2003) provided a follow-up analysis of the Amrein and Berlinger (2002) study using a comparison group of 14-18 states, depending on the year of the NAEP exam, that did not attach sanctions to high-stakes testing. Table 5 represents the assessments, grades, and years used in Rosenshine's study. Rosenshine (2003) concluded, "increases in the *clear* high-stakes states were much higher than the increases in comparison states" (p. 2). Rosenshine (2003) reported the following effect sizes based on comparisons: .35 for fourth grade Mathematics, .79 for eighth grade Mathematics, and .61 for fourth grade Reading. These effect sizes are considered moderate to large (Rosenshine, 2003). However, Rosenshine (2003) reported that the increases in scores

were not consistently reported across all states with high-stakes testing policies. For example, South Carolina, Massachusetts, and Alabama did well in fourth grade Mathematics while New Mexico, West Virginia or Kentucky did not. Additionally, Indiana and Alabama did well in eighth grade Mathematics while New Mexico and Missouri did not. In fourth grade Reading, Louisiana, Delaware, and Virginia were high performing while Missouri and New Mexico were not. Perhaps these discrepant scores require additional research regarding the potential mediating effect of internal accountability mechanisms within external accountability policies. Specifically, states with lackluster gains should investigate whether their state assessment policies are aligned to the national policies with regard to assessment content and rigor.

Table 5

Data Sources for Rosenshine Study

Assessment	Data Source	Years
NAEP	Grade 4 Mathematics composite scores	1996-2000
NAEP	Grade 8 Mathematics composite scores	1996-2000
NAEP	Grade 4 Reading composite scores	1994-1998

Note. Assessments used by Rosenshine from "High-Stakes Testing: Another Analysis" by B. Rosenshine, 2003, *Education Policy Analysis Archives, 11*(24).

Summary of Accountability in Education

The term accountability in education is expansive and refers to both the internal accountability of schools and external accountability imposed upon schools. Researchers report that schools with strong internal accountability systems will likely perform better within a system of external accountability (Abelmann et al., 1999; Elmore, 2005; Fuhrman, 1999; Newmann, King, & Rigdon, 1997). Even though no single definition of

accountability exists, both Kirst (1990) and Darling-Hammond (1989) present potential typologies that serve to explicate external accountability through potential mechanisms. The design and purpose of the present study aligns with specific mechanisms outlined by Darling-Hammond (1989). Darling-Hammond (1989) presents five accountability mechanisms or sources that influence public education including: political, legal, bureaucratic, market-based systems, and professional. Of the five sources, No Child Left Behind conveys a strong bureaucratic system of external accountability by imposing consequences or rewards over schools based on student performance on high-stakes tests.

NCLB policies have been in place for nearly a decade, however the research regarding influences of external accountability over school improvement has returned mixed results (Lee, 2006). With the exception of national studies using low-stakes tests (i.e., Amrein & Berlinger, 2002; Carnoy & Loeb, 2002; Hanushek & Raymond, 2005; Rosenshine, 2003), few studies exist that assess the influences of current NCLB accountability policies over school improvement. The present study will add to the existing scholarship regarding accountability and public education by quantifying the improvement of schools, operating in a system of external accountability, using value-added methodologies.

Value-Added Modeling

Central to this study's thesis is the following question: will increased accountability of schools to external agents improve the performance of schools? The public demands greater accountability over schools, yet researchers grapple with fundamental questions such as; what is school improvement? And, can the improvement of schools be measured with reliability and validity (Raudenbush, 2004)? These

questions and others like them guide the existing literature pertaining to the school's effect on student learning.

As researchers study the effects of schools on student outcomes, findings are often met with debate regarding a range of related topics, including: the type of outcome variables (i.e., high-stakes tests versus authentic assessments), costs of monitoring teachers and schools (i.e., teacher autonomy versus mandated test-prep), and the validness of methodology used to separate the school's effect on student learning, apart from variables beyond the school's control (i.e., student background and school context; Linn, 2008; Raudenbush, 2004; Wiliam, 2010). The ongoing psychometrical discourse in the existing literature is sufficiently deliberated by those with a full command of assessment quality and reliability. Therefore, recalling that the aim of this study is to explore the influence of external accountability policies on school improvement efforts, the following section will evade the debate regarding suitability of outcome variables and direct-costs on teacher autonomy while focusing on the aptness of Pennsylvania's Value-Added Assessment System (PVAAS) as a measure of school improvement. To that end, this section will (1) develop a theory-base for value-added modeling (VAM), (2) define how VAM differs from traditional approaches to gauge school improvement, and (3) critique specific studies regarding PVAAS as an acceptable method of VAM.

Theory of Value-Added Modeling (VAM)

Underlying the current, outcome-based system of accountability is the assumption that student performance can provide a reasonable estimate of school performance (Harris, 2010). Philosophically, VAM espouses a similar, outcome-based approach to determine school effectiveness. Agreement exists among many researchers that value-

added comparisons are an improvement over traditional, mean proficiency score comparisons to determine the effectiveness of schools (Hershberg et al., 2004; Linn, 2008; Meyer, 2000; Raudenbush, 2004; Sanders, 2000). However, agreement becomes less evident as researchers debate the suitability of one value-added model over another to isolate the effects of the school apart from contributing covariates (i.e., school context and student characteristics). Therefore, to provide both a theoretical base for employing VAM in education accountability as well its relevance in the present study, this section will illustrate VAM conceptually, and then specifically as it relates to PVAAS. To begin this depiction, this study refers to the work of Willms and Raudenbush (1989) who describe VAM through two distinct "effect-types."

According to Willms and Raudenbush (1989), VAM incorporates the following components to determine the extent schools influence student outcomes: outcomes (*Y*), school practices (*P*), school context (*C*), student background (*S*), and random error (*e*). School practices *P* include a range of factors that are controlled at the school level such as leadership, curriculum, resource use, and instructional techniques. School context *C* refers to the social, economic, and demographic factors that are beyond the control of the school. The composition of a student's background *S* includes his or her aptitude, race, gender, prior schooling, and socioeconomic status. Even though school practice *P* and school context *C* may be perceived to be highly correlated variables, they are distinctly separate. School context *C* will interact at varying degrees with both the school practice *P* and student characteristics *S*. The interaction between *P, C,* and *S* is a source of ongoing deliberation with regards to employing one value-added model over another. This study assumes that outcomes *Y* are a function of the school practice *P*, school

context *C*, and student background *S* at varying degrees and dependent upon the intended effect type.

Type A effect. Raudenbush and Willms (1995) define Type A effects as the difference between an individual's actual outcome and the individual's expected outcome in a "typical school." Type A effects could be hypothesized in an experimental design by randomly assigning students with similar *S* to different schools *j*.

$$A_{ij} = P_{ij} + C_{ij}$$

Type A effects conclude that the value-added of one school over another would be the function of both school practice *P* and school context *C*. Even though parents may look more closely at the Type A effect when selecting a school (school context is arguably an attractive component of schools), policymakers and district administrators would be unable to, fairly, determine the effectiveness of a school apart from the school context using this model.

Type B effect. Type B effects intentionally isolate the effect of school practice *P* on student outcomes *Y* by controlling the factors associated with school context *C* and student background *S*. To hypothesize a Type B effect using an imaginary experiment, consider a population of students with identical backgrounds *S,* who are assigned to two clusters of schools, *J* and *J',* both clusters share the same school context *C* but employ distinctly different practices *P* and *P*'. Ideally, *J* would represent the cluster of schools employing typical or average practices *P* while *J'* would represent the school employing the experimental practice *P'*. In this design, all things are considered equal except the practices of the school. Therefore, any change in outcome *Y* could infer a causal link to school practice *P*.

$$B_{ij} = P_{ij}$$

According to Raudenbush and Willms (1995) Type B effects could be used to determine the value a particular school adds to a student's outcomes *Y*.

Raudenbush (2004) concluded that the Type B effects necessary for assessing the school improvement apart from the context of school and student are not "plausibly detectable from accountability data alone" (p. 12). However, Sanders (2000) contests that improvement effects can be fairly estimated at both the school and the teacher level using the existing data from criterion referenced tests (i.e., Mathematics and Reading PSSA). Sanders and Horn (1998) contend that the "layered" (LA) value-added model employed by PVAAS, where the effect of prior schooling and teacher contributions remains with the student as they move from grade to grade, can sufficiently isolate both school and teacher effects. The LA model assumes, generally, that students maintain all of the knowledge they learned in previous grades (Ballou, Sanders, & Wright, 2004). Consider the following for the LA model for student *i* in grade *g:*

$$y_{ig} = \mu_g + \sum \Psi_{ig}\theta_g + \varepsilon_{ig}$$

Consider the following layered years of data where μ represents the school mean in grade *g* and $\Psi_{ig}\theta_g$ represents the interaction of factors related to practice and context on student *i* in grade *g:*

$$y_{i3} = \mu_3 + \Psi_{i3}\theta_3 + \varepsilon_{i3}$$

$$y_{i4} = \mu_4 + \Psi_{i3}\theta_3 + \Psi_{i4}\theta_4 + \varepsilon_{i4}$$

$$y_{i5} = \mu_5 + \Psi_{i3}\theta_3 + \Psi_{i4}\theta_4 + \Psi_{i5}\theta_5 + \varepsilon_{i5}$$

$$y_{i6} = \mu_6 + \Psi_{i3}\theta_3 + \Psi_{i4}\theta_4 + \Psi_{i5}\theta_5 + \Psi_{i6}\theta_6 + \varepsilon_{i6}$$

$$y_{i7} = \mu_7 + \Psi_{i3}\theta_3 + \Psi_{i4}\theta_4 + \Psi_{i5}\theta_5 + \Psi_{i6}\theta_6 + \Psi_{i7}\theta_7 + \varepsilon_{i7}$$

Pennsylvania does not calculate teacher effects for the purpose of evaluation. Therefore, the PVAAS methodology will differ slightly from the general LA model referenced above.

Traditional Model Versus VAM

Traditional, cross-sectional practices evaluate the effectiveness of schools by comparing the mean percentage of students scoring proficient on high-stakes tests to other schools as well as established AYP proficiency targets. In this same manner, school improvement is determined by the difference from previous years' proficiency percentages to the current year. However, Raudenbush (2004) reported "in current accountability systems, student intake and instructional effectiveness are confounded by some unknown degree, calling into question any inferences about school effectiveness from these data" (p. 7). Other researchers report that the traditional cross-sectional method is incapable of sufficiently controlling for the student-level variables (i.e., socioeconomic status and prior achievement) known through research to have a close relationship with student performance on high-stakes tests (Hershberg, Simon, & Lea-Kruger, 2004; Linn, 2008; Meyer, 2000; Olsen, 2007; Sanders, 2000). VAM addresses the limitations presented by No Child Left Behind's (NCLB), traditional, cross-sectional status approach by using individual student assessment scores (longitudinally), rather than a school-wide, cross-sectional proficiency percentage to determine growth.

Pennsylvania Value-Added Assessment System (PVAAS)

Pioneered by William Sanders in the 1980s, VAM became institutionalized by Tennessee's department of education throughout the 1990s and is now largely known as the "Sander's Model" or Tennessee's Value Added Assessment System (TVAAS;

Sanders & Horn, 1998). In Pennsylvania, the State Board of Education passed an initial resolution that adopted the TVAAS methodology and thus established the Pennsylvania Value-Added Assessment System (PVAAS) in 2002 (PVAAS: Statewide Plan, 2006). Throughout this section TVAAS and PVAAS will denote essentially the same method for VAM and used interchangeably.

Pennsylvania's timeline for employing PVAAS included three pilot years (Pilot 1- 2002, Pilot 2- 2004 and Pilot 3- 2005). By 2007, all Pennsylvania school districts received PVAAS comprehensive reports regarding student achievement. Beginning with the assessments administered in 2010, standardized, PVAAS Average Growth Indices (AGIs) were publicized for the purposes of comparing schools. From the onset, Pennsylvania established two clear goals for PVAAS including: (1) to employ PVAAS as a tool for the continuous improvement of schools and (2) to include "growth model" as calculated through PVAAS, to the existing status measures for achieving AYP (PVAAS: Statewide Plan, 2006). This study will discuss the increasing role PVAAS plays in Pennsylvania's school improvement process by determining and comparing the "value-added" of high schools.

PVAAS uses a mixed-model methodology to develop a longitudinal analysis of student scores from high-stakes tests (Sanders & Horn, 1998). Essentially, PVAAS applies the LA method of using multiple years of assessment scores to model each student's pattern of learning. Through layers of assessments and grades, PVAAS determines if each student's learning deviates from what would be considered "normal" based on each student's longitudinal testing pattern. If the deviation reflects positively, then causal inferences can be made regarding the school's added value. A negative

deviation would infer the opposite. The causal inferences can only be assumed, based on the control PVAAS has over confounding variables at the school and student level. Pennsylvania aggregates student-level, value-added gains (+ or -) at the school-level to determine the Average Gain Index (AGI). The AGI can be used to compare schools at the state level.

Univariate response model. PVAAS employs several statistical models depending on the needs of the analysis and the availability of data (Wright, White, Sanders & Rivers, 2010). For example, PVAAS reports the value-added of a school in terms of residual gains using the multivariate response model (MRM). MRM is a multivariate, longitudinal, linear mixed model. PVAAS employs MRM when using scaled data from students assessed in subsequent grades (i.e., from grade four to five). For Pennsylvania high schools, when subsequent grades are not assessed, PVAAS uses a univariate response model (URM) to make projections from grade eight to grade eleven. Essentially, PVAAS can project an eighth grade student's score to a future eleventh grade assessment using a composite representation of prior Mathematics and Reading assessment scores weighted by the appropriateness of the assessment on the intended projection. Subsequently, the projection can be used to determine if the student's predicted score on the eleventh grade assessment is more or less than the actual score to determine the value-added (+ or -) of the school with regard to that student.

Consider the projection for i^{th} student using URM employed by PVAAS:

$$C_i = \hat{\mu}_y + \hat{\beta}_1(x_{i1} - \hat{\mu}_1) + \hat{\beta}_2(x_{i2} - \hat{\mu}_2) + \cdots$$

The projection *C* denotes a one-number composite representing the accumulation of all available (historical) assessment data on student *i*. The categorical variable, *1,2,3…n*,

represents the school the student attended for each subject, grade, and year of the response variable *y*. The regression coefficients, $\hat{\beta}_n$, represent the various weighted properties applied to each assessment x_n. Therefore, PVAAS addresses problems plaguing both Type A and Type B effects through layered, longitudinal composites at the student level.

Limitations of PVAAS. A review of the current literature regarding PVAAS, as well as other value-added models, revealed a potential limitation to the present study. PVAAS omits student-level covariates within its application (McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004). VAM models generally consider factors as fixed or random. Fixed factors are those factors such as gender or ethnicity that will not change or expand in the layering of the statistical model. In contrast, many factors at the school and student level (i.e., school-wide demographics or SES) will change or expand within the model and are handled through the computation as random effects. The PVAAS methodology ascribes to a mixed-model where some factors are fixed and other random. This poses concerns from researchers, since PVAAS does not include student-level covariates, known to correlate with achievement, within each layer of the model. Essentially, these variables are controlled, implicitly, within the longitudinal design of the model (Ballou, Sanders, & Wright, 2004). Researchers defending the approach employed by PVAAS to control covariates, generally do so on the following grounds:

> Because [demographic and SES variables] are correlated with otherwise unmeasured variation in school and teacher quality, the coefficients on these variables will capture part of what researchers are trying to measure with

> residuals. Predictors of school and teacher effectiveness will accordingly be biased toward zero. (Ballou, Sanders, & Wright, 2004, pp. 38-39)

However, many within the research community disagree with this rationale (Lissitz & Doran, 2009).

Research on Value-Added Modeling

McCaffrey et al. (2004) studied the influence of covariates across four generic value-added models to estimate teacher effects on student outcomes. Although McCaffrey et al. (2004) were examining teacher effects, rather than school effects, the issues regarding the handling of covariates are similar within each model (Ballou, Sanders, & Wright, 2004). Moreover, McCaffrey et al. (2004) specifically addressed the LA model, employed by PVAAS, and therefore brings particular importance to the present study.

To evaluate each of the generic models and make comparisons between models, McCaffrey and colleagues (2004) proposed a general model of VAM and confirmed that each of the generic models represented a restricted manifestation of their general model. Quite literally the general model became the framework to study the effects of covariates within and between the various generic models. As indicated previously, McCaffrey et al. addressed issues arising from the exclusion of covariates from the LA model by studying the manner by which omitted covariates bias parameter estimates when students are stratified at the classroom-level by those covariates. Essentially, if the same student demographics that are omitted by the model (i.e., SES) contribute to student-classroom assignment (including for the purpose of heterogeneity), then the omitted covariates would become confounded with teacher effects. In these cases, student covariates would

likely become confounded with school effects as well. This issue is of paramount concern especially when using VAM for the purpose of teacher evaluation and evaluating school improvement.

Notwithstanding the limitations of VAM estimations to assess teacher quality, McCaffrey et al. (2004) acknowledge the work of other researchers who contend that omitting correlates may be appropriate in some cases and, in fact, consistent with the findings from their research. Essentially, McCaffrey et al. reported that the issue of including or omitting student-level covariates was "complex and dependent upon the distribution of the omitted covariates and the assignment of students to teachers" (p. 95).

Ballou, Sanders, & Wright (2004) evaluated the method PVAAS employs for controlling covariates using the existing TVAAS database. In this study, Ballou and colleagues used Mathematics, Reading, and language arts assessments to question the extent a teacher's effect would differ if student characteristics and school context factors were included in the model. For comparison, the following student-level characteristics were entered into the TVAAS system: (1) SES, (2) race other than white, and (3) gender. To control for school context effect, known through this study as "peer effects," the aggregate composition of SES, was included in the model at both the school and the class-level. In sum, three variations of the adjusted model were compared to the unadjusted model. These variations include: (1) student level-covariates and teacher (by year) fixed effects, (2) student-level covariates and constrained teacher effects with SES aggregated at the school level, and (3) student-level covariates and constrained teacher effects with SES aggregated at the grade (within school) level.

Ballou, Sanders, & Wright (2004) found that the inclusion of student-level demographic covariates (variation 1) in the adjusted model had little effect on the unadjusted TVAAS results for all subjects. Agreement between variation 1 and the unadjusted model was 2.7 times more likely in Reading, 3.5 times more likely in language arts and 8.5 times more likely in Mathematics. Thus the study concluded that controlling for student characteristics, only, had little effect on TVAAS. Conversely, when controlling for SES, aggregated at either the school or the grade-within-school level (variation 2 and 3 respectfully), their findings were more substantial. However, due to the high standard errors resulting in the second and third variations, their study concluded that including the aggregate SES at both the school and grade-within-school results were suspect and unstable.

These researchers offered four possible explanations why adding student-level covariates (in the adjusted model) had little impact on the estimation of teacher effects. First, Ballou, Sanders, & Wright (2004) considered the possibility that student characteristics (SES and race) were evenly distributed across classes and thus resulted in no significant change in teacher-effects. However, the researchers reported a mean SES proportion of .47 with a standard deviation of .23 and a mean race (non-white) proportion of .54 with a standard deviation of .21. These reported proportions and accompanying standard deviations describe considerable variability of student characteristics across teachers rendering this possible explanation as, not likely. The second possible explanation from Ballou, Sanders, & Wright (2004) suggested that the effect of student variables is not large enough to make a meaningful difference in estimated teacher effects using either model. However, the researchers assessed the correlation between the

adjusted models and a fixed-effect model (with no controls for student characteristics) and found inconsistencies between the models while also reporting a high correlation between the adjusted and unadjusted TVAAS model. Therefore, this explanation lacks plausibility while supporting the research that SES and other student demographics are correlated with student performance. Third, the correlation between the adjusted and unadjusted models is a result of shrinkage. Finally, SES and demographic covariates do not add enough to the model beyond the covariance found in assessment scores. Essentially, the performance of each student over time is collected in the scores of each assessment and therefore included, implicitly within the model.

Both studies shine a light on a limitation of PVAAS as well as implications for the present study using Pennsylvania high schools. For example, McCaffrey et al. (2004) supported Ballou, Sanders, & Wright (2004) in that student-level covariates should not be controlled within the model if students are *not* stratified across classes. Simply, unlike elementary schools, high school students change classes frequently during the day, throughout the week and over a school year. Even though a high school student's demographics will plausibly become a component within grades, based on other factors such as level of classes (i.e., AP versus vocational setting), students are not typically stratified (intentionally) across classes to the extent they might be in an elementary school. Therefore, optimism exists that the PVAAS model, though imperfect, is a more reliable means of measuring school improvement than existing models and thus compatible with the goals and design of this study.

Summary of Value-Added Modeling

Value-Added Modeling (VAM) in education was developed to more accurately report the effectiveness of schools with regard to student outcomes. In the 1990's the Tennessee Department of Education, under the direction of William Sanders, enhanced the use of VAM for use in public education (Sanders & Horn, 1998). Currently, the Tennessee Value Added Assessment System (TVAAS) is largely credited with bringing VAM to public education. Pennsylvania adopted the TVAAS methodology for it own use by employing the Pennsylvania Value-Added Assessment System (PVAAS) in 2002 (PVAAS: Statewide Plan, 2006).

PVAAS applies a mixed-model methodology and longitudinal analysis of student test scores to make predictions on future high-stakes tests (Sanders & Horn, 1998). For high schools in Pennsylvania, PVAAS projects a student's score on eleventh grade assessments using a combination of all prior assessment scores (grades eight and below). The projection can then be used to determine if the student's predicted score on the eleventh grade assessment is more or less than his or her actual score to determine the value-added (+ or -) of the school on that student's performance. At the school level, student scores are averaged to create the overall Growth Measure (GM) for each school. The GM divided by the standard error produces the *standardized* Average Growth Index (AGI), used to compare all high schools in Pennsylvania.

The analysis of studies regarding value-added assessment models, including those employed by PVAAS, reveals how this new methodology is undergoing rigorous scrutiny by experts within the field of education as well as the greater scientific community. At the present time, no studies exist within the corpus of related literature that suggest the

statistical approaches employed by PVAAS are a detriment to the current accountability policies. Moreover, the analysis of literature regarding VAM, gives promise that the methodology employed by PVAAS to determine school improvement, albeit imperfect, is an improvement over the cross-sectional status models (Hershberg, Simon, & Lea-Kruger, 2004; Linn, 2008; Meyer, 2000; Olsen, 2007; Sanders, 2000).

CHAPTER 3

Method

No Child Left Behind (NCLB) identifies schools as needing improvement through the accountability structure known as Adequate Yearly Progress (AYP). Schools that continually fail to meet yearly AYP targets are sanctioned at increasingly higher levels until they are ultimately restructured or taken over by the state. However, researchers have reported that the traditional approach to identify schools for improvement, using "mean proficiency scores," is imprecise and fails to isolate the performance of schools over time or apart from the covariates associated with achievement (Davies, 2008; Forte, 2010; Harris, 2010; Hershberg, Simon, & Lea-Kruger, 2004; Linn, 2008; Meyer, 2000; Raudenbush, 2004). Therefore, school improvement and the overall effectiveness of NCLB cannot be determined without a system to accurately measure school performance.

The Pennsylvania Value Added Assessment System (PVAAS) uses longitudinal student data and a predictive methodology to determine the growth of schools on high-stakes tests. Though imperfect, the methodology used by PVAAS is reported as a more accurate measure of school performance by controlling the covariates that influence achievement (Harris, 2010, Reeves, 2005; Sanders, 2000). Hence, PVAAS adds to the overall purpose of this study, which is to determine the overall utility of external accountability in public education.

Research Questions

This study used scaffolded research questions in conjunction with research hypotheses to advance the overall purpose of this study. First, this study sought to determine whether PVAAS, as a statistical instrument, successfully quantified the

improvement of schools across the School Improvement Classifications (SIC). Second, to the extent that PVAAS could quantify and demonstrate a significant relationship across SICs, schools identified as Making Progress would need to demonstrate higher gains than schools Needing Improvement to infer a plausible relationship between external accountability and school improvement. The third and final step to this analysis was to identify the ongoing improvement of schools by assessing the progress of schools identified as Making Progress in subsequent testing years.

The present study seeks answers to the following research questions using the 11th grade proficiency exams.

Question 1a: Does the PVAAS Average Growth Index (AGI) in Mathematics differ between Pennsylvania high schools across School Improvement Classifications (SIC)?

Question 1b: Does the PVAAS Average Growth Index (AGI) in Reading differ between Pennsylvania high schools across School Improvement Classifications (SIC)?

Question 2a: Do high schools identified as Making Progress demonstrate higher PVAAS Average Growth Indices (AGI) in Mathematics than high schools Needing Improvement?

Question 2b: Do high schools identified as Making Progress demonstrate higher PVAAS Average Growth Indices (AGI) in Reading than high schools Needing Improvement?

Question 3a: Do the PVAAS Average Growth Index (AGI) gains in Mathematics for high schools identified as Making Progress continue to subsequent testing years?

Question 3b: Do the PVAAS Average Growth Index (AGI) gains in Reading for high schools identified as Making Progress continue to subsequent testing years?

Hypotheses

According to Fraenkel and Wallen (2009), restating research questions as hypotheses can be advantageous when the purpose of the research is to both build a body of knowledge and answer specific questions. Fraenkel and Wallen contend that including research hypotheses allows the researcher "to make specific predictions based on prior evidence or theoretical argument" (p. 46). The present study investigates the following hypotheses regarding the role and relevance of a value-added model (specifically PVAAS) as a tool for advancing educational accountability under the NCLB mandate.

Hypothesis 1a: There will be identifiable and statistically significant differences in high schools' Mathematics PVAAS Average Growth Indices (AGI) as a function of School Improvement Classification (SIC).

Hypothesis 1b: There will be identifiable and statistically significant differences in high schools' Reading PVAAS Average Growth Indices (AGI) as a function of School Improvement Classification (SIC).

Hypothesis 2a: Schools identified as Making Progress in 2010 will have higher PVAAS Mathematics Average Growth Indices (AGI) than schools identified as Needing Improvement.

Hypothesis 2b: Schools identified as Making Progress in 2010 will have higher PVAAS Reading Average Growth Indices (AGI) than schools identified as Needing Improvement.

Hypothesis 3a: Schools Making Progress in 2010 will continue to demonstrate similar PVAAS gains in Mathematics during 2011.

Hypothesis 3b: Schools Making Progress in 2010 will continue to demonstrate similar PVAAS gains in Reading during 2011.

Participant Schools

The present study examined the 11th grade PVAAS AGI scores from the existing population of Pennsylvania public high schools for academic year 2009-2010. High schools were those schools that met all of the following criteria: (1) had a graduating class during the 2010 school year, (2) had PVAAS AGI scores reported for both Mathematics and Reading reflecting only the eleventh grade during the 2010 assessment year, (3) did not contain grade eight or below, (4) was not a private school, (5) was not a charter school, (6) was not a school operated by an intermediate unit, (7) was not a vocational-technical school, (8) was not state-owned, (9) was not a special program school, (10) was not a juvenile detention school, (11) was not a private academy, (12) was not an approved private school, and (13) had at least two years of prior eleventh grade assessment data (to omit schools that were restructured and newly appointed schools). 426 high schools met the criteria above and became the sample for this study. This sample serves as a snapshot of all Pennsylvania high schools, past and future.

Research Design

Views differ on how best to classify non-experimental research and no single structure for classifying such studies is broadly endorsed by the research community (Fraenkel & Wallen, 2009). Many educational researchers would arguably classify the present, non-experimental, study as causal-comparative or ex post facto, due to the use of both categorical (independent) and quantitative (dependent) variables. However, causal-comparative or ex post facto research intuitively suggests that the researcher has identified a problem as well as potential factors that may associate with causation (Ary, Jacobs & Sorensen, 2010; Fraenkel & Wallen, 2009). This assertion is inexact with regard to the aim of the present study that seeks to understand and quantify the extent the *existing* system of accountability is influencing schools *already* identified at pre-determined levels of school improvement.

Johnson (2001) introduced a typology that classifies all non-experimental research across two dimensions, (1) time and (2) research objective. This study uses historical data (time) to explain the extent a possible relationship already exists (research objective). Using Johnson's typology, this study is classified as "retrospective, explanatory" non-experimental research. Table 6 represents the classification system proposed by Johnson for all non-experimental research.

Table 6

Johnson's Typology for Non-Experimental Research

Research objective	Retrospective	Cross-sectional	Longitudinal
Descriptive	Retrospective, descriptive study (Type 1)	Cross-sectional, descriptive study (Type 2)	Longitudinal, descriptive study (Type 3)
Predictive	Retrospective, predictive study (Type 4)	Cross-sectional, predictive study (Type 5)	Longitudinal, predictive study (Type 6)
Explanatory	Retrospective, explanatory study (Type 7)	Cross-sectional, explanatory study (Type 8)	Longitudinal, explanatory study (Type 9)

Note. Adapted from "Toward a New Classification of Non-experimental Quantitative Research" by B. Johnson, 2001, *Educational Researcher, 30*(2), p.10.

Dependent Variable

The results from the 2010 eleventh grade Mathematics and Reading PSSAs were used to quantify the dependent variable as the PVAAS Average Growth Index (AGI) for Pennsylvania high schools. The Average Growth Index is the Growth Measure (see definition of terms in Chapter One) for the 2010 assessment year divided by its standard error (PVAAS Statewide Core Team for PDE, 2011). PVAAS calculated the Mathematics AGI and Reading AGI for each high school represented in the study, separately.

Independent Variable

The independent variable was a modified version of the 2010 No Child Left Behind (NCLB) School Improvement status for high schools in Pennsylvania. Modifications to the NCLB status were made at two levels (SIC A and SIC B) as represented in Figure 1.

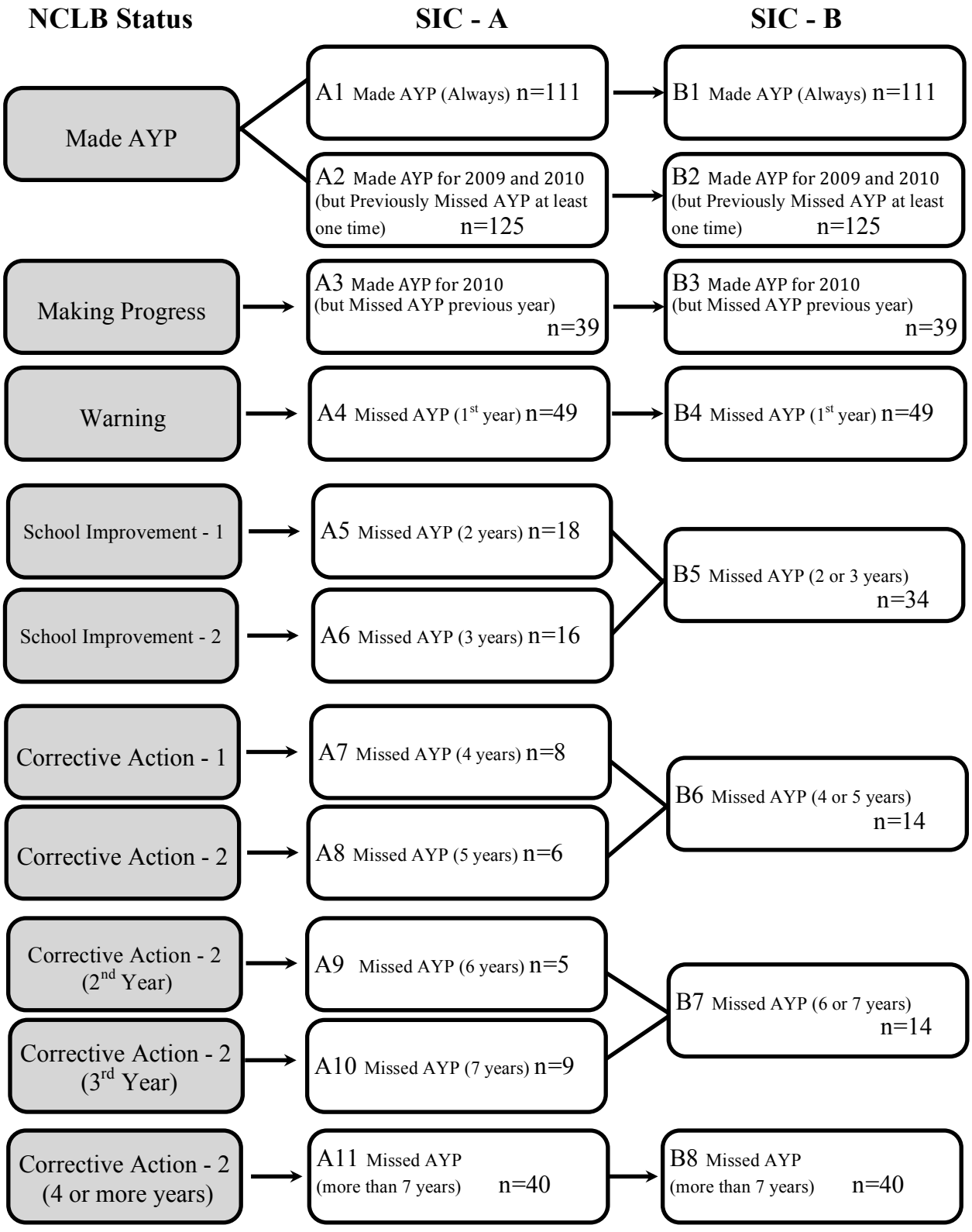

Figure 1. Continuum of Adjustments to the Independent Variable from NCLB Status to School Improvement Classifications for SIC A and SIC B; *n*= number of schools in the group

The first, SIC A, level of adjustment divided the Made AYP status into two separate School Improvement Classifications (SIC). SIC A1 represents the group of high schools that "Always Made AYP since 2003," the year NCLB began reporting the status of schools. SIC A2 represents the group of schools that Made AYP in 2009 and 2010, but previously Missed AYP (at least one time) in the years from 2003-2008. The only other adjustment during the first level of modification was to combine all high schools that Missed AYP for more than seven years into one group, SIC A11.

Subsequent to the SIC A level of modification, the group size (i.e., the number of high schools within a classification) was very small for the following levels: A7, A8, A9, and A10; group size was marginal for A5 and A6. Thus, a second level of modification was performed to increase the size of these six smallest groups by pairing "like" groups. The second level of modification reduced the eleven SIC A levels to eight SIC B levels. SIC B levels were used for all statistical tests.

Procedures for Retrieving Data from 2010 Testing

A single table in the form of an Excel spreadsheet was used to record the following data for this study: "District Name," "School Name," "State School ID," "2010 PVAAS Math," "2010 PVAAS Reading," "2010 School Improvement Classification (SIC) code," and "2010 NCLB School Improvement Status."

Average Growth Index. The dependent variable (high school PVAAS AGI scores) was retrieved July 20, 2011 from the Pennsylvania Department of Education PVAAS Public Reporting website at: https://pvaas.sas.com/evaas/signin.jsf
Since a data file could not be downloaded, each high school's 2010 PVAAS AGI scores for both Mathematics and Reading were transcribed, manually, into corresponding fields

in the aforementioned Excel spreadsheet. The complete 2010 data set for 426 high schools is located in Appendix D.

School Improvement Classification. On August 1, 2011, the independent variable, 2010 School Improvement Classification (SIC), was recorded by first downloading the "Schooldata_Part1" data file from the Pennsylvania Department of Education Website at: http://paayp.emetric.net/Content/datafiles/2010AYP_SchoolData_Part1.CSV. The downloaded data file was delimited to include only existing high school records by filtering on the following criteria: (1) include assessment scores for grade eleven, (2) exclude assessment scores for grades eight and below, (3) include only schools with a graduating class, and (4) include only public schools.

Additional Data Retrieved Regarding 2011 Testing

To determine the improvement of schools over time, the 2011 PVAAS AGI scores for Mathematics and Reading were retrieved on October 26, 2011 for the 39 schools representing the 2010 Making Progress group from the Pennsylvania Department of Education PVAAS Public Reporting website at: https://pvaas.sas.com/evaas/signin.jsf. Additionally, the School Improvement Classification (SIC) was recorded by downloading and recording the NCLB status from the "Schooldata_Part1" data file from the Pennsylvania Department of Education Website at: http://paayp.emetric.net/Content/datafiles/2011AYP_SchoolData_Part1.CSV. Modifications to the NCLB status for these schools were made in accordance to the process defined by Figure 1. The 2011 data for 39 schools classified as Making Progress

in 2010 were added to the original spreadsheet under the following column headings: "2011 SIC B Level," "2011 Math PVAAS AGI," and "2011 Reading PVAAS AGI."

Statistical Analyses

Tests for Hypothesis 1. One-Way Analysis of Variance (ANOVA) was conducted for both Mathematics and Reading scores to test for significant differences in PVAAS AGI scores across the levels of the independent variable (School Improvement classification). Effect size was calculated in terms of Omega-square (ω^2).

Tests for Hypothesis 2. The Tukey test was used (post hoc) to determine whether the 2010 PVAAS AGI scores for high schools identified as Making Progress were significantly higher than the scores for high schools Needing Improvement. The effect size was calculated in terms of Cohen's d.

Tests for Hypothesis 3. A paired-samples t-test was conducted to compare the Reading and Mathematics gains from the 2010 Making Progress group to the gains from the same cohort of high schools in 2011.

CHAPTER 4

Results

The purpose of this chapter is to report the results of the statistical analyses conducted for this study. All data were collected from publicly available sources and organized in Excel spreadsheets. The software program, Statistical Package for Social Sciences (SPSS; version 18) was used to analyze all results. The analyses were conducted for each hypothesis and the results are reported by content area, Mathematics then Reading.

The reporting of results begins with a set of preliminary analyses provided for purposes of general background information. Here the reader will find a histogram showing the overall frequency distribution of PVAAS scores (for Mathematics and subsequently for Reading) for the entire set of 426 high schools. Here one also finds summary statistics for the 11 SIC A school improvement classifications and for the 8 SIC B classifications. [The definitions for each of the SIC A and SIC B classifications, along with corresponding NCLB categories, can be found in Figure 1.] These preliminary analyses serve three functions. First, the histograms convey important information about the range of scores and shape of the frequency distribution with regard to the dependent variable. Second, the summary statistics demonstrate that the results are much the same whether one uses the SIC A or SIC B classification of the independent variable. Lastly, the preliminary analyses provide transparency and justification for the decision to test all hypotheses using the SIC B classification system.

The main statistical analyses are then reported as they relate to each of the stated research hypotheses. The dependent variable is the PVAAS AGI score (for Mathematics

or Reading) for each high school; the independent variable is the SIC B school improvement category that the school falls in. For the convenience of the reader, the SIC B codes and descriptions are presented in Table 7.

Table 7

Classification of the Independent Variable as SIC B

SIC Code	SIC Description	NCLB School Improvement Status
B1	Made AYP (Always)	Made AYP
B2	Made AYP for 2009 and 2010 (but Previously Missed AYP at least one time)	Made AYP
B3	Made AYP for 2010 (but Missed AYP previous year)	Making Progress
B4	Missed AYP (1st year)	Warning
B5	Missed AYP (2 or 3 years)	School Improvement 1 & 2
B6	Missed AYP (4 or 5 years)	Corrective Action 1 & 2 (1st year)
B7	Missed AYP (6 or 7 years)	Corrective Action 2 (2nd & 3rd year)
B8	Missed AYP (more that 7 years)	Corrective Action 2 (4 or more years)

Note. SIC= School Improvement Classification; SIC codes B4 to B8= "Needing Improvement."

Findings for Mathematics

The frequency distribution of the 2010 PVAAS AGI scores for Mathematics is represented in Figure 2 as a histogram. The distribution is approximately normal with a slight positive skew. The full frequency distribution table for Mathematics scores appears in Appendix B.

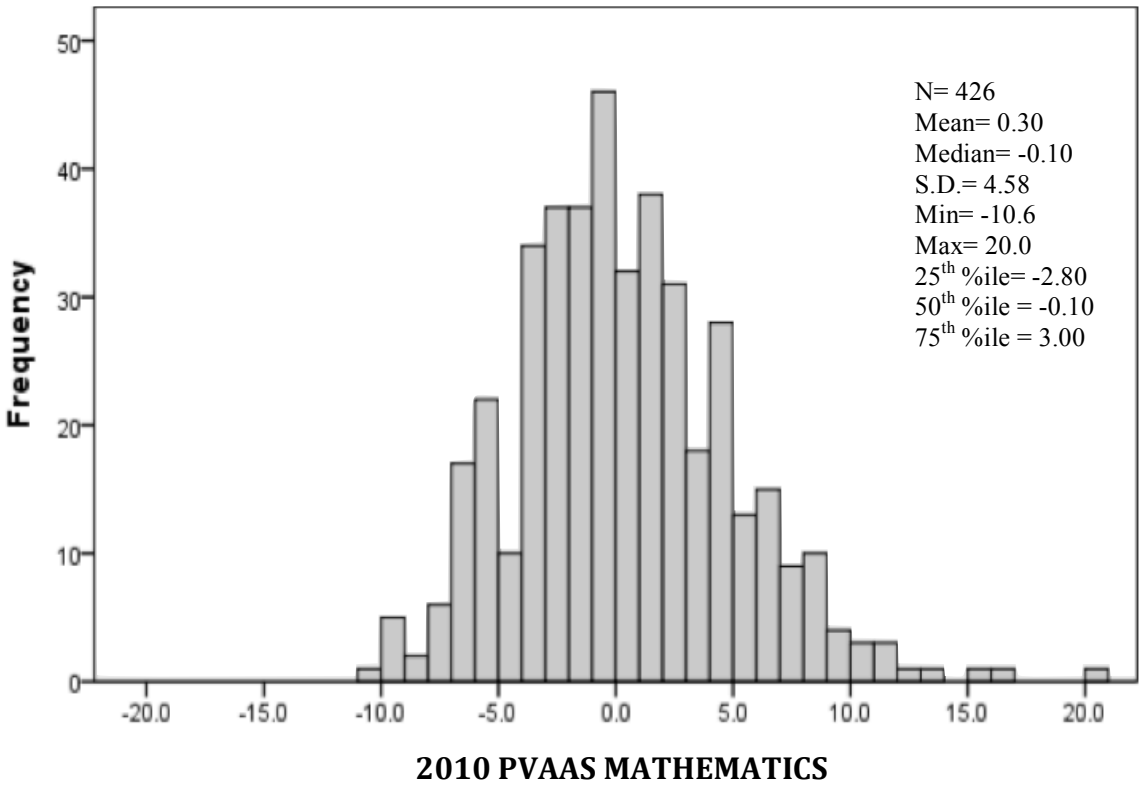

Figure 2. Frequency Histogram for 2010 PVAAS AGI scores in Mathematics

Table 8 reports summary statistics for Mathematics by the first modification to the independent variable known as School Improvement Classification (SIC A). The percentile equivalent for each group was included to represent each group's relative position within the overall distribution.

It should be noted that sample size was very small for groups A7-A10 and marginal for groups A5 and A6.

Table 8

Summary Information for Mathematics (SIC-A)

SIC Code	*n*	*Mean*	*(%ile Equiv.)*	*SD*	*SEmean*	*95% CI [LL, UL]*
A1	111	-0.19	(49.5)	4.07	0.39	[-0.96, 0.57]
A2	125	1.17	(61.1)	4.32	0.39	[0.40, 1.93]
A3	39	2.80	(74.4)	4.51	0.72	[1.33, 4.26]
A4	49	-0.26	(48.5)	4.66	0.67	[-1.60, 1.08]
A5	18	0.13	(52.8)	4.09	0.97	[-1.90, 2.17]
A6	16	-0.73	(43.2)	4.32	1.08	[-3.03, 1.57]
A7	8	0.53	(54.3)	4.08	1.44	[-2.89, 3.94]
A8	6	1.15	(60.9)	6.55	2.68	[-5.73, 8.03]
A9	5	2.08	(68.9)	7.67	3.43	[-7.44, 11.60]
A10	9	-1.62	(35.2)	5.51	1.84	[-5.86, 2.61]
A11	40	-2.22	(30.6)	4.77	0.75	[-3.75, -0.70]
Total	426	0.30	-	4.58	0.22	[-0.13, 0.74]

Note. SIC= School Improvement Classification; %ile Equiv= Percentile Equivalent for each mean; SD= standard deviation; SEmean= standard error of the mean; CI= confidence interval; LL= lower limit; UL= upper limit

Figure 3 depicts the mean and confidence interval for each SIC A classification. As evidenced by the confidence bands, three of the eleven groups (SIC A2, A3, and A11) had means significantly different from zero. The means for A2 (Made AYP for 2009 and 2010, but Missed AYP in a prior year) and A3 (Making Progress 2010) were significantly above zero. The mean for A11 (the group that Never Made AYP) was significantly below zero.

Confidence intervals for groups A7-A10 are inordinately wide, due in large part to the small sample size in these conditions.

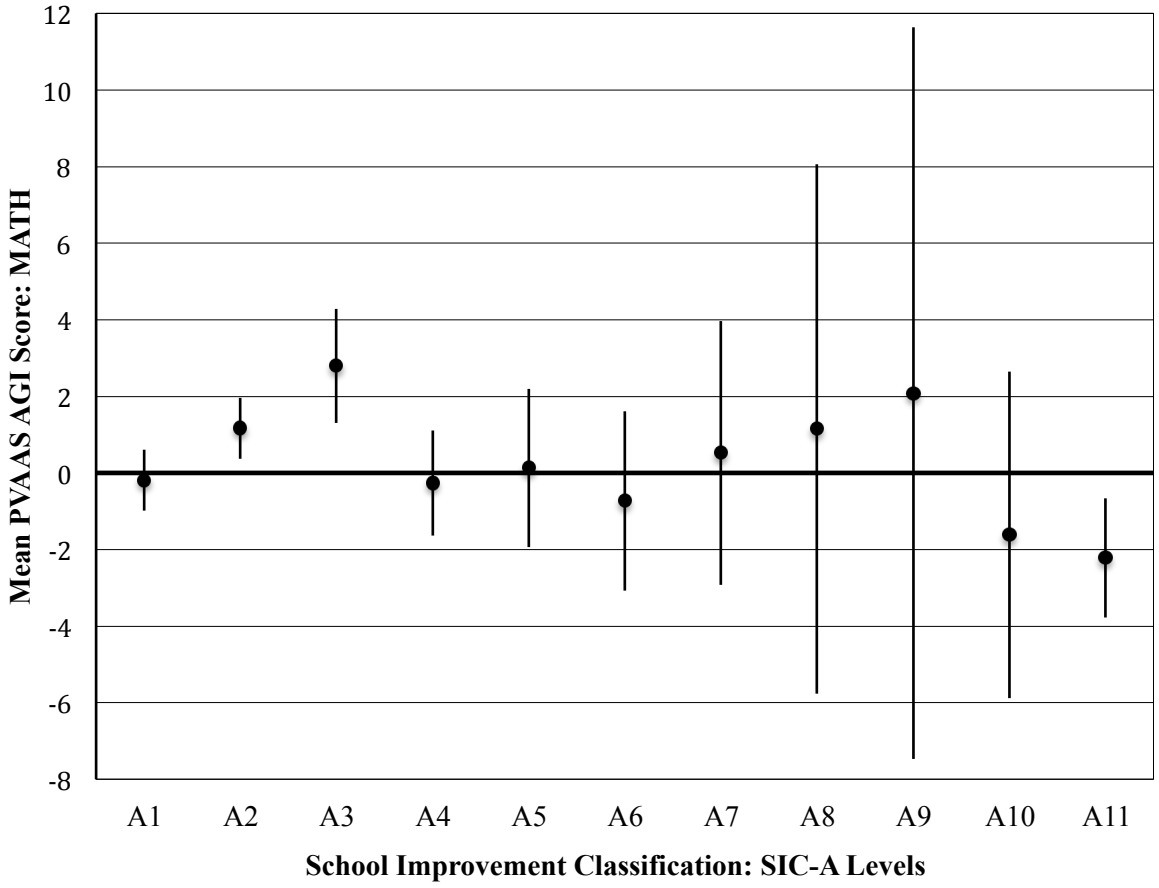

Figure 3. Mean and 95% Confidence Intervals by School Improvement Classification (A) for Mathematics

Table 9 reports the summary statistics for SIC B classification of the independent variable. Sample sizes for SIC B classifications (B6 and B7) are an improvement over their SIC A counterparts (A7-A10), though they continue to be small relative to the other B classifications.

Table 9

Summary Information for Mathematics (SIC-B)

SIC Code	*n*	*Mean*	*(%ile Equiv.)*	*SD*	*SEmean*	*95% CI [LL, UL]*
B1	111	-0.19	(49.5)	4.07	0.39	[-0.96, 0.57]
B2	125	1.17	(61.1)	4.32	0.39	[0.40, 1.93]
B3	39	2.80	(74.4)	4.51	0.72	[1.33, 4.26]
B4	49	-0.26	(48.5)	4.66	0.67	[-1.60, 1.08]
B5	34	-0.27	(48.0)	4.16	0.71	[-1.73, 1.18]
B6	14	0.79	(56.7)	5.06	1.35	[-2.13, 3.71]
B7	14	-0.3	(47.4)	6.33	1.69	[-3.96, 3.36]
B8	40	-2.22	(30.6)	4.77	0.75	[-3.75, -0.70]
Total	426	0.30	-	4.58	0.22	[-0.13, 0.74]

Note. SIC= School Improvement Classification; %ile Equiv= Percentile Equivalent; SD= standard deviation; SEmean= standard error of the mean; CI= confidence interval; LL= lower limit; UL= upper limit.

Figure 4 depicts the mean Mathematics AGI score along with the corresponding confidence interval for each SIC B classification. Means for groups B2 and B3 were significantly above zero; the mean for B8 was significantly below zero. [These are the same groups previously labeled SIC A2, A3, and A11.]

The confidence intervals for the two smallest groups (SIC B6 and B7) are a bit wide but are an improvement over SIC A.

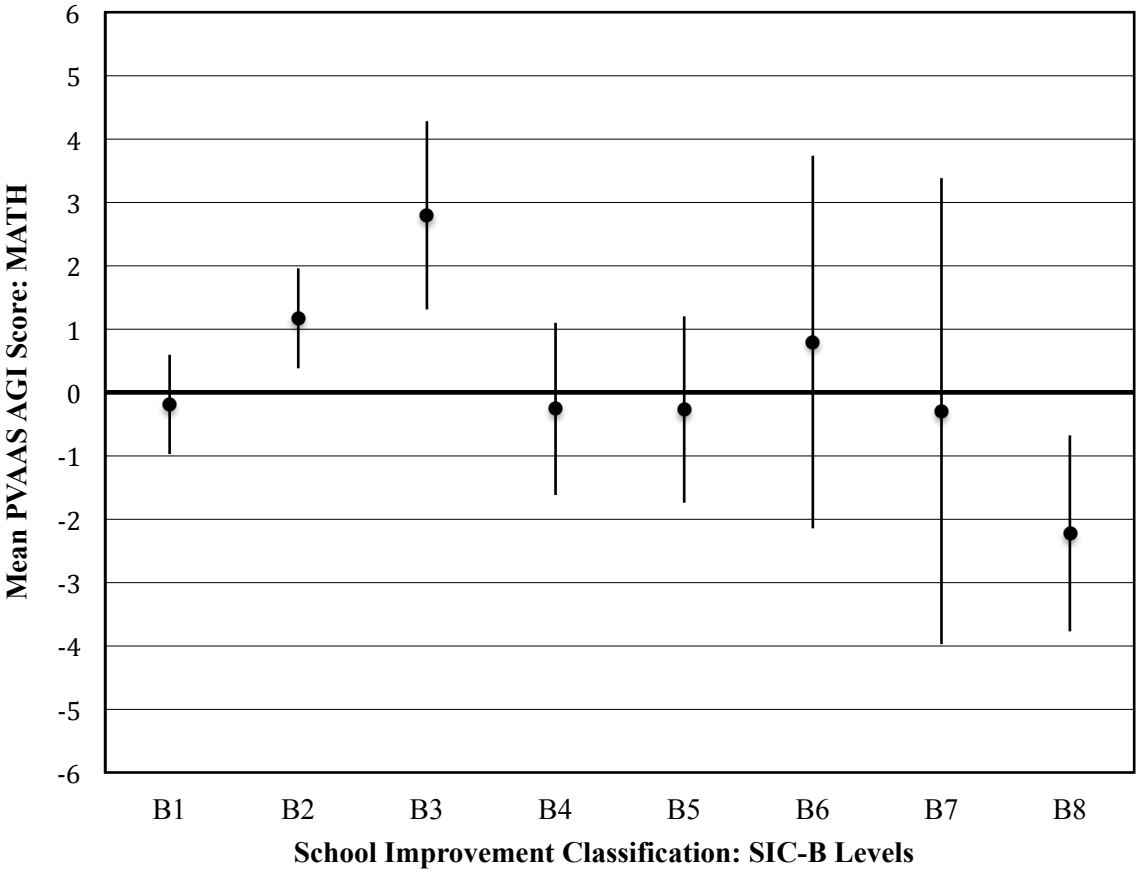

Figure 4. Mean and 95% Confidence Intervals by School Improvement Classification (B) for Mathematics

Analysis for Hypothesis One

Hypothesis 1: There will be identifiable and statistically significant differences in high schools' Mathematics PVAAS Average Growth Indices (AGI) as a function of School Improvement Classification (SIC).

A one-way analysis of variance (ANOVA) was conducted with PVAAS AGI Mathematics scores as the dependent variable and School Improvement Classification (SIC B) as the independent variable (grouping variable). The ANOVA summary table is provided in Table 10. ANOVA results revealed that mean AGI scores differed significantly among school improvement groups, $F(7, 418) = 4.72$, $p<.001$.

Effect size as measured by omega-square ($\omega^2=.06$) was small. School Improvement Classification (SIC B) accounted for about 6% of the total variability in Mathematics PVAAS AGI scores.

Table 10

ANOVA Summary Table for Mathematics

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	652.79	7	93.26	4.72	.000
Within Groups	8260.05	418	19.76		
Total	8912.83	425			

Analysis for Hypothesis Two

Hypothesis 2: Schools identified as Making Progress in 2010 will have higher PVAAS Mathematics Average Growth Indices (AGI) than schools identified as Needing Improvement.

Post-hoc analysis was performed using the Tukey procedure to determine whether schools identified as Making Progress (SIC B3) outperformed schools Needing Improvement (SIC B4-B8). Table 11 reports all pairwise mean differences in Mathematics. In total, four significant differences were found.

Table 11

Tukey Post hoc Comparisons for Mathematics

SIC	B1	B2	B3	B4	B5	B6	B7	B8
B1	---	1.36	2.99**	-0.07	-0.08	0.98	-0.11	-2.03
B2	-1.36	---	1.63	-1.42	-1.44	-0.37	-1.47	-3.39**
B3	-2.99**	-1.63	---	-3.05*	-3.07	-2.00	-3.09	-5.02**
B4	0.07	1.42	3.05*	---	-0.02	1.05	-0.04	-1.97
B5	0.08	1.44	3.07	0.02	---	1.07	-0.03	-1.95
B6	-0.98	0.37	2.00	-1.05	-1.07	---	-1.09	-3.02
B7	0.11	1.47	3.09	0.04	0.03	1.09	---	-1.92
B8	2.03	3.39**	5.02**	1.97	1.95	3.02	1.92	---

Note. * indicates p<.05. ** indicates p<.01

Table 12 provides a summary of the significant differences. Congruent with the second hypothesis, three of the four significant findings involved the Making Progress group (SIC B3) scoring higher than other groups. In addition to outperforming two groups of schools Needing Improvement (SIC B4 and B8), the Making Progress group also scored higher than the group of schools identified as "Always Making AYP" (SIC

B1). The fourth significant finding for Mathematics involved the group identified as "Making AYP for 2009 and 2010" (SIC B2) outperforming schools that "Missed AYP for more than 7 Years" (SIC B8).

Table 12 also reports Cohen's $\hat{d}$ as measure of effect size for each significant pairwise comparison. Effect size was medium to large in all instances.

Table 12

Summary of Significant Findings for Mathematics

Significant Pairwise Comparisons			Difference Between Means	Cohen's $\hat{d}$
B3	Greater than	B1	2.99**	.67
		B4	3.05*	.69
		B8	5.02**	1.13
B2	Greater than	B8	3.39**	.76

Note. * indicates $p < .05$, ** indicates $p < .01$

Analysis for Hypothesis Three

Hypothesis 3: The schools identified as Making Progress in 2010 will continue to demonstrate similar PVAAS gains in Mathematics during 2011.

A paired-samples t-test, as reported in Table 13, compared the Mathematics PVAAS scores from the 2010 Making Progress group to the scores from the same cohort of high schools in 2011. In 2010, there were 39 schools in the Making Progress group. Two of these schools did not have PVAAS scores reported in 2011 and were dropped from the analysis.

The 2010 Making Progress group had a mean Mathematics score of +3.00 (SD=4.54) on the 2010 assessment. A year later, the mean Mathematics score for these schools fell to 1.70 (SD=4.43). The difference was statistically significant at the .05 alpha level, t(36)=2.17, $p = 0.04$. Effect size was small, Cohen's $\hat{d}$= .29.

Table 13

Paired Sample t-test for 2010 Making Progress Group for Mathematics

		Paired Differences							
					95% CI				Sig.
	N	M_D	SD_D	SE_D	LL	UL	t	*df*	(2-tailed)
Pair: Math 2010 for MP 2010-Math 2011 for MP 2010	37	1.31	3.65	0.60	0.09	2.52	2.17	36	.036*

Note. M_D= mean of the difference; SD_D= standard deviation of the difference; SE_D= standard error of the difference; CI= confidence interval; LL= lower limit; UL= upper limit; df= degrees of freedom; indicates $p < .05$

It is important to note that the third research hypothesis forecasted that the gains made by the Making Progress group in 2010 would be achieved again in 2011 (an indication of ongoing school improvement). Instead, PVAAS Mathematics scores dropped significantly in that timeframe.

Further Examination of Data Relevant to Hypothesis 3

Table 14, Table 15, and Table 16 provide data relevant to Hypothesis 3 for each of the 39 high schools classified Making Progress in 2010. Table 14 reports the characteristics for the 13 schools (of the aforementioned 39) that Made AYP in 2011. Table 15 reports the same information for the 24 schools that Missed AYP in 2011. Table 16 reports characteristics for the two schools that did not report PVAAS scores in 2011. These tables, in addition to providing Mathematics scores, also include Science and Writing scores that will be addressed in Chapter Five as an area for further research.

Table 14

School Characteristics for the 2010 Making Progress Schools that Made AYP in 2011: Mathematics

		PVAAS AGI			*PVAAS AGI*	
		Mathematics			*Science*	*Writing*
School ID	*2009 NCLB Status*	*2010*	*2011*	*2011-2010*	*2011*	*2011*
HS 319	S. I.- 1	8.1	6.3	-1.8	2.7	0.9
HS 318	C. A. -2 (1 year)	8.0	9.7	1.7	2.1	4.0
HS 312	S. I.- 2	4.5	2.9	-1.6	6.2	5.1
HS 307	C. A.- 1	2.7	1.4	-1.3	3.2	-0.5
HS 306	S. I.- 1	2.4	3.7	1.3	4.4	-4.9
HS 302	S. I.- 2	1.6	0.1	-1.5	2.2	1.6
HS 303	S. I.- 1	1.6	-0.4	-2.0	0.9	4.0
HS 300	C. A.- 1	1.0	3.2	2.2	0.6	3.9
HS 296	S. I.- 1	-0.1	-2.1	-2.0	-0.1	-1.8
HS 291	S. I.- 1	-1.7	4.5	6.2	3.2	-3.8
HS 290	S. I.- 1	-2.0	-2.9	-0.9	-2.2	-1.5
HS 289	S. I.- 1	-2.4	-1.5	0.9	0.6	0.6
HS 288	S. I.- 1	-3.9	-3.9	0.0	1.3	-1.8
Total		1.52	1.62	0.09	1.93	0.45

Note: n=13. S. I. -1= School Improvement 1; S . I. -2= School Improvement 2; C. A. 1= Corrective Action 1; C. A. -2 (1 year)= Corrective Action 2 (1st year). 2011-2010= difference between AGI scores from 2011 to 2010

Table 15

School Characteristics for the 2010 Making Progress Schools that Missed AYP in 2011: Mathematics

		PVAAS AGI			*PVAAS AGI*	
		Mathematics			*Science*	*Writing*
School ID	*2009 NCLB Status*	*2010*	*2011*	*2011-2010*	*2011*	*2011*
HS 324	C. A. -2 (1 year)	12.0	3.4	-8.6	1.1	1.9
HS 323	C. A. -2 (4+)	11.9	9.1	-2.8	-8.4	-1.5
HS 321	S. I.- 1	8.8	6.6	-2.2	8.7	13.1
HS 322	S. I.- 2	8.8	10.7	1.9	7.4	1.4
HS 320	C.A.2 (2nd year)	8.7	9.4	0.7	0.4	1.0
HS 317	S. I.- 1	7.5	4.9	-2.6	11.1	1.3
HS 316	C.A. 1	6.7	-1.5	-8.2	2.1	2.5
HS 315	S. I.- 2	6.5	6.1	-0.4	3.5	7.9
HS 314	S. I.- 1	6.3	1.3	-5.0	2.2	-6.9
HS 313	S. I.- 1	6.1	4.0	-2.1	-5.0	5.9
HS 311	C.A. 1	4.4	-0.8	-5.2	0.1	-0.3
HS 310	C.A.2 (2nd year)	4.2	-0.4	-4.6	-3.0	0.3
HS 309	S. I.- 1	3.6	-0.8	-4.4	-5.7	-2.6
HS 308	C.A.2 (2nd year)	3.2	1.3	-1.9	-9.6	-2.5
HS 304	S. I.- 2	1.8	-4.7	-6.5	-3.7	-4.6
HS 305	S. I.- 1	1.8	-0.6	-2.4	-1.6	-3.8
HS 301	S. I.- 1	1.5	1.8	0.3	-2.9	-4.6
HS 299	S. I.- 1	0.7	0.8	0.1	-3.1	-2.0
HS 298	S. I.- 1	0.4	-3.3	-3.7	-0.1	-11.1
HS 297	S. I.- 2	-0.1	6.9	7.0	3.6	-2.6
HS 295	C.A. 1	-0.6	-6.9	-6.3	1.9	-9.1
HS 292	S. I.- 2	-1.5	-3.8	-2.3	-5.2	-5.2
HS 287	C.A. 1	-4.3	2.5	6.8	-0.6	0.0
HS 286	S. I.- 1	-7.1	-4.2	2.9	-0.6	-1.8
	Total	3.80	1.74	-2.06	-0.31	-0.97

Note: n= 24. The 2010 Making Progress schools that Missed AYP in 2011 would fall one level below their 2009 NCLB Status (see Figure 1 for current NCLB status for these schools). S. I. -1= School Improvement 1; S. I. -2= School Improvement 2; C. A. 1= Corrective Action 1; C. A. -2 (1 year)= Corrective Action 2 (1st year); C. A. -2 (2 year)= Corrective Action 2 (2nd year); C. A. -2 (4+ year)= Corrective Action 2 (4 or more years). 2011-2010= difference between AGI scores from 2011 to 2010.

Table 16

2010 Making Progress Schools without PVAAS Results Reported in 2011: Mathematics

		PVAAS AGI Mathematics			PVAAS AGI Science	PVAAS AGI Writing
School ID	*2009 NCLB Status*	*2010*	*2011*	*2011-2010*	*2011*	*2011*
HS 294	-	-0.9	-	-	-	-
HS 293	S. I.- 1	-1.2	-	-	-	-
	Total	-1.05	-	-	-	-

Note: n=2. HS 294 was closed in June 2011. HS 293 did not Make AYP during 2011. The 2010 Making Progress schools that Missed AYP in 2011 would fall to one level below their 2009 NCLB Status (see Figure 1 for current NCLB status for these schools). S. I. -1= School Improvement 1.

The line graph in Figure 5 depicts the 2010 and 2011 mean Mathematics AGI scores for the subgroup of 13 schools that Made AYP in 2011 (see Table 14) and the subgroup of 24 schools that Missed AYP in 2011 (see Table 15), from the group of 37 schools classified as Making Progress in 2010 that had data reported for both years.

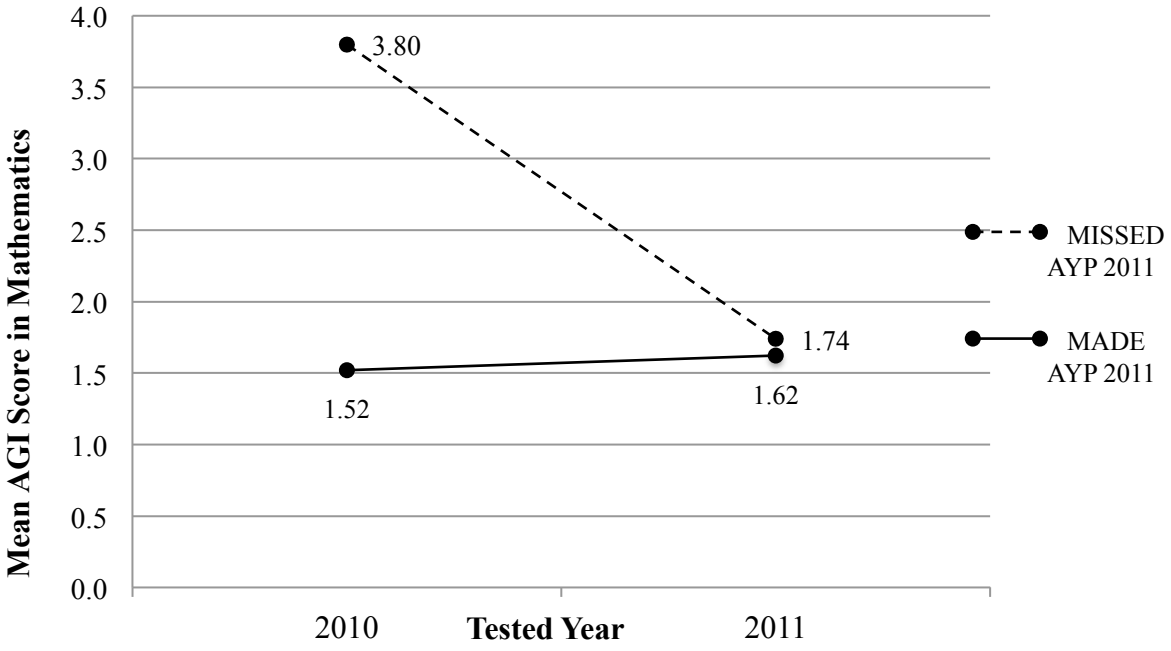

Figure 5. Mean AGI scores in Mathematics for 2010 and 2011, by schools that Made AYP and those that Missed AYP during the 2011 school year (restricted to schools classified as Making Progress in 2010)

The following points should be noted with regard to Figure 5: (1) The subgroup that Made Progress in 2010 and Made AYP in 2011 sustained their 2010 PVAAS Mathematics gains into 2011. (2) The subgroup that Made Progress in 2010 but Missed AYP in 2011 showed considerable decline in 2011. (3) The subgroup that Missed AYP in 2011 started off with the higher mean in 2010, while the subgroup that Made AYP in 2011 had the lower mean.

Findings for Reading

Figure 6 shows the frequency distribution of the 2010 PVAAS AGI scores for Reading as a histogram. The distribution is approximately normal with a slight positive skew, due in large part to the seven schools that had AGI scores of +15.0. These outliers will be addressed later in the report. The full frequency distribution table for Reading scores is provided in Appendix C.

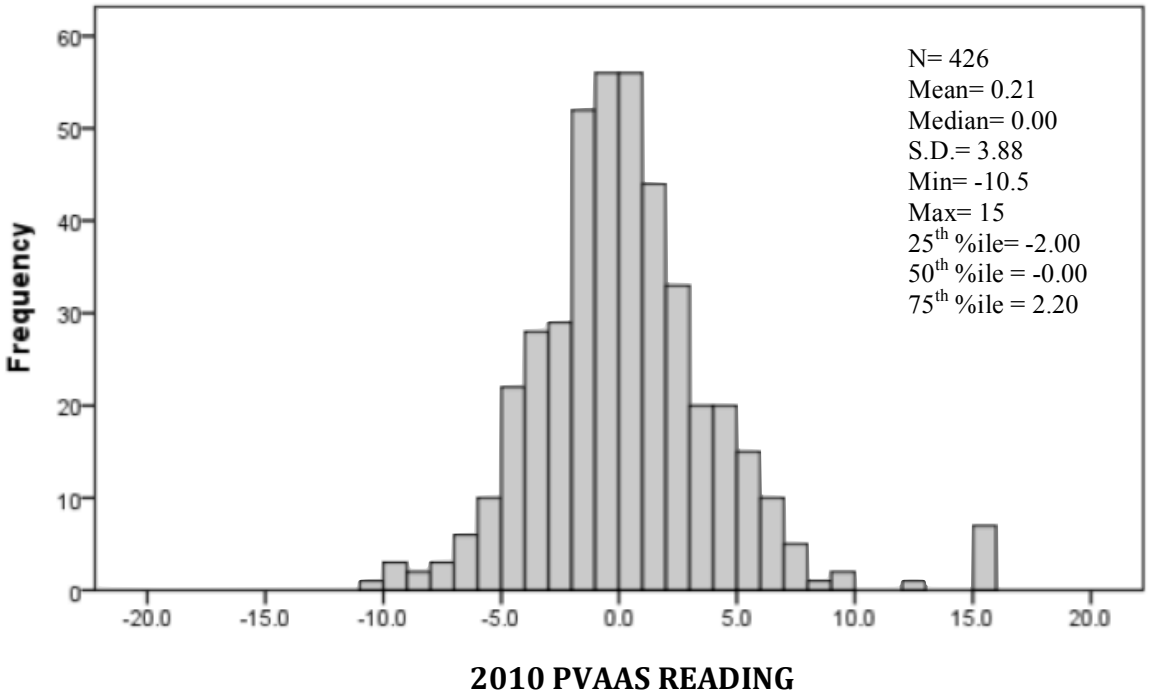

Figure 6. Frequency Histogram for 2010 PVAAS AGI Scores in Reading

Table 17 reports summary statistics for Reading by the first modification to the independent variable referred to in this study as SIC A. Group sizes are identical to those reported previously for Mathematics under the SIC A classification (see Table 8).

Table 17

Summary Information for Reading (SIC-A)

SIC Code	*N*	*Mean*	*(%ile Equiv.)*	*SD*	*SEmean*	*95% CI [LL, UL]*
A1	111	0.08	(52.0)	2.86	0.27	[-0.46, 0.61]
A2	125	0.44	(57.2)	3.15	0.28	[-0.12, 1.00]
A3	39	3.51	(84.8)	6.51	1.04	[1.40, 5.62]
A4	49	-0.34	(47.1)	3.42	0.49	[-1.32, 0.65]
A5	18	-0.13	(49.5)	3.70	0.87	[-1.97, 1.71]
A6	16	-0.81	(41.2)	4.18	1.04	[-3.03, 1.42]
A7	8	2.53	(79.1)	3.78	1.34	[-0.64, 5.69]
A8	6	0.28	(54.6)	2.79	1.14	[-2.65, 3.22]
A9	5	0.54	(58.6)	3.03	1.36	[-3.23, 4.31]
A10	9	-0.68	(43.1)	3.41	1.14	[-3.30, 1.94]
A11	40	-2.42	(22.0)	3.69	0.58	[-3.60, -1.24]
Total	426	0.21	-	3.88	0.19	[-0.16, 0.58]

Note. SIC= School Improvement Classification; %ile Equiv= Percentile Equivalent; SD= standard deviation; SEmean= standard error of the mean; CI= confidence interval; LL= lower limit; UL= upper limit

Figure 7 depicts the mean and confidence interval for each SIC A classification. As evidenced by the confidence bands, the mean for A3 (Making Progress) was significantly above zero; the mean for A11 (the group that never Made AYP) was significantly below zero.

Confidence intervals for groups A5-A10 are wide, due in large part to the small sample sizes for these groups. The confidence interval for SIC A3 (Making Progress) is also wide. This, in part, is due to the fact that all seven schools with the outlier score of +15.0 (the highest score in the distribution) fell in this category.

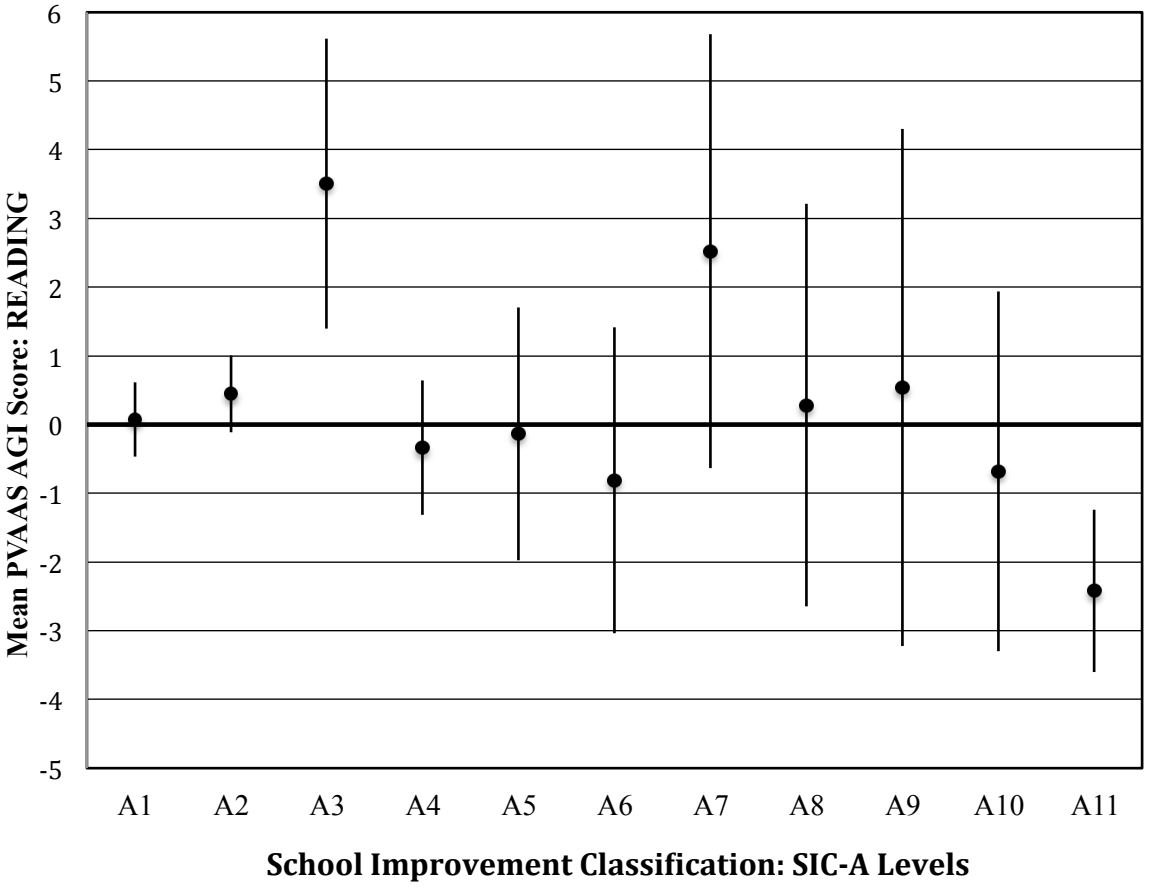

Figure 7. Mean and 95% Confidence Intervals by School Improvement Classification Code (A) for Reading

Table 18 reports the summary statistics for SIC B classification of the independent variable. The SIC B classifications have previously been defined in Figure 1 and Table 7.

Table 18

Summary Information for Reading (SIC-B)

SIC Code	*N*	*Mean*	*(%ile Equiv.)*	*SD*	*SEmean*	*95% CI [LL, UL]*
B1	111	0.08	(52.0)	2.86	0.27	[-0.46, 0.61]
B2	125	0.44	(57.2)	3.15	0.28	[-0.12, 1.00]
B3	39	3.51	(84.8)	6.51	1.04	[1.40, 5.62]
B4	49	-0.34	(47.1)	3.42	0.49	[-1.32, 0.65]
B5	34	-0.45	(46.5)	3.89	0.67	[-1.81, 0.91]
B6	14	1.56	(69.9)	3.47	0.93	[-0.44, 3.57]
B7	14	-0.24	(48.2)	3.21	0.86	[-2.10, 1.61]
B8	40	-2.42	(21.8)	3.69	0.58	[-3.60, -1.24]
Total	426	0.21	-	3.88	0.19	[-0.16, 0.58]

Note. SIC= School Improvement Classification; %ile Equiv= Percentile Equivalent; SD= standard deviation; SEmean= standard error of the mean; CI= confidence interval; LL= lower limit; UL= upper limit.

Figure 8 depicts the mean AGI score for Reading with corresponding confidence interval for each SIC B classification. The mean for group B3 was significantly above zero. The mean for B8 was significantly below zero. [These are the same groups labeled previously SIC A3 and A11.]

The confidence intervals for the two smallest groups (SIC B6 and B7) are a noticeable improvement over SIC A.

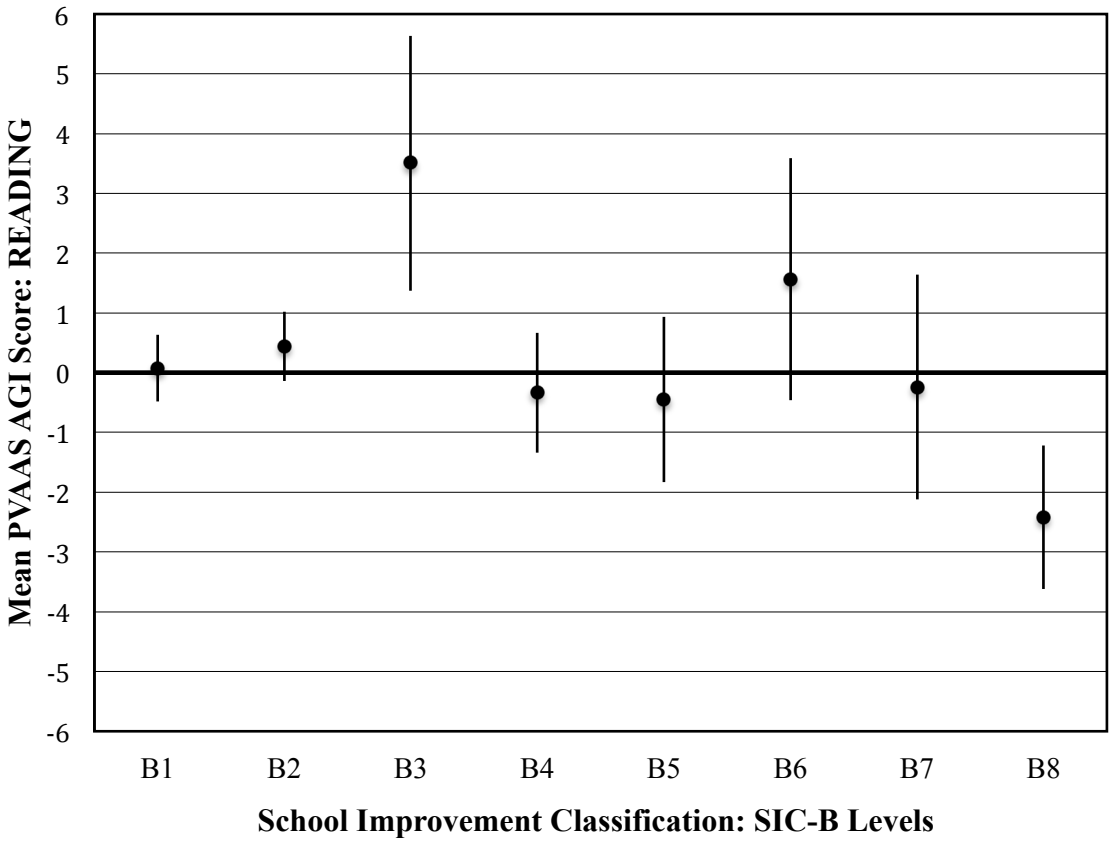

Figure 8. Mean and 95% Confidence Intervals by School Improvement Classification Code (B) for Reading

Analysis for Hypothesis One

Hypothesis 1: There will be identifiable and statistically significant differences in high schools' Reading PVAAS Average Growth Indices (AGI) as a function of School Improvement Classification (SIC).

A one-way analysis of variance (ANOVA) was conducted with PVAAS AGI Reading scores as the dependent variable and School Improvement Classification (SIC B) as the independent variable. The ANOVA summary table is provided in Table 19. The analysis revealed that mean AGI scores differed significantly among School Improvement Classifications (SIC B levels), $F(7, 418) = 8.15$, $p<.001$.

Effect size as measured by omega-square ($\omega^2=.11$) was small to medium. School Improvement Classification (SIC B) accounted for about 11% of the total variability in Reading PVAAS AGI scores.

Table 19

ANOVA Summary Table for Reading

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	767.51	7	109.65	8.15	.000
Within Groups	5622.73	418	13.45		
Total	6390.24	425			

Analysis for Hypothesis Two

Hypothesis 2: Schools identified as Making Progress in 2010 will have higher PVAAS Reading Average Growth Indices (AGI) than schools identified as Needing Improvement.

Post-hoc analysis using the Tukey procedure was performed to determine whether schools identified as Making Progress (SIC B3) outperformed schools Needing Improvement (B4-B8) in the content area of Reading. Table 20 reports all pairwise mean differences. In total, nine significant differences were found.

Table 20

Tukey Post hoc Comparisons for Reading

SIC	B1	B2	B3	B4	B5	B6	B7	B8
B1	---	0.37	3.43**	-0.41	-0.52	1.49	-0.32	-2.49**
B2	-0.37	---	3.06**	-0.78	-0.89	1.12	-0.69	-2.86**
B3	-3.43**	-3.06**	---	-3.84**	-3.96**	-1.94	-3.75*	-5.93**
B4	0.41	0.78	3.84**	---	-0.12	1.90	0.09	-2.09
B5	0.52	0.89	3.96**	0.12	---	2.01	0.21	-1.97
B6	-1.49	-1.12	1.94	-1.90	-2.01	---	-1.81	-3.98*
B7	0.32	0.69	3.75*	-0.09	-0.21	1.81	---	-2.18
B8	2.49**	2.86**	5.93**	2.09	1.97	3.98*	2.18	---

Note. * indicates p<.05. ** indicates p<.01

Table 21 provides a summary of the significant differences. Six of the nine significant findings involved the 2010 Making Progress group (SIC B3). The Making Progress group outperformed every group with one exception, that being the group comprised of schools that "Missed AYP for 4 or 5 years" (SIC B6). Not only did group

B3 significantly outperform four of the five groups Needing Improvement (SIC B4, B5, B7, and B8), it also outperformed two groups that Made AYP in 2010 (SIC B1 and B2). The other significant findings involved the group that "Missed AYP for more than 7 years" (SIC B8). This group scored lower than four groups (one of which was B3, as mentioned previously).

Table 21 also reports Cohen's $\hat{d}$ as a measure of effect size for each significant pairwise comparison. Effect sizes for all comparisons involving SIC B3 were large. The effect sizes for comparisons involving SIC B8 were medium to large.

Table 21

Summary of Significant Findings for Reading

Significant Pairwise Comparisons			Difference Between Means	Cohen's $\hat{d}$
B3	Greater than	B1	3.43**	.93
		B2	3.06*	.83
		B4	3.84**	1.05
		B5	3.96**	1.08
		B7	3.75*	1.02
		B8	5.93**	1.61
B8	Less than	B1	2.49**	.68
		B2	2.86**	.78
		B3	5.93**	1.61
		B6	3.98*	1.08

Note. * indicates $p < .05$, ** indicates $p < .01$

Analysis for Hypothesis Three

Hypothesis 3: The schools identified as Making Progress in 2010 will continue to demonstrate similar PVAAS gains in Reading during 2011.

A paired-samples t-test, as reported in Table 22, compared the Reading PVAAS scores from the 2010 Making Progress group to the scores from the same cohort of high schools in 2011. Two of the original 39 schools did not have PVAAS scores reported in 2011 and thus were dropped from the analysis.

The 2010 Making Progress group had a mean Reading score of +3.39 (SD=6.30) on the 2010 assessment. A year later, the mean Reading score for these schools fell to +0.34 (SD=3.29). The difference was statistically significant at the .01 alpha level, $t(36)=3.01$, $p = 0.005$. Effect size was medium, Cohen's $\hat{d}= .61$.

Table 22

Paired Sample t-test for 2010 Making Progress Group for Reading

		Paired Differences							
					95% CI				Sig.
	N	M_D	SD_D	SE_D	LL	UL	T	*df*	(2-tailed)
Pair: Reading 2010 for MP 2010-Reading 2011 for MP 2010	37	3.05	6.16	1.01	1.00	5.11	3.01	36	.005**

Note. M_D= mean of the difference; SD_D= standard deviation of the difference; SE_D= standard error of the difference; CI= confidence interval; LL= lower limit; UL= upper limit; df= degrees of freedom;
** indicates $p < .01$

Hypothesis 3 predicted that the gains made by the Making Progress group in 2010 would be achieved again in 2011. Instead 2010 PVAAS Reading scores dropped significantly in 2011.

Further Examination of Data Relevant to Hypothesis 3

Table 23, Table 24, and Table 25 provide data relevant to Hypothesis 3 for each of the 39 high schools classified as Making Progress in 2010. Table 23 reports the characteristics for the 13 schools (of the original 39) that Made AYP in 2011. Table 24 reports the same information for the 24 schools that Missed AYP in 2011. Table 25 shows characteristics for the two schools that did not report PVAAS scores in 2011. These tables also provide Science and Writing scores that will be addressed in Chapter Five.

Table 23

School Characteristics for the 2010 Making Progress Schools that Made AYP in 2011: Reading

		PVAAS AGI			*PVAAS AGI*	
		Reading			*Science*	*Writing*
School ID	*2009 NCLB Status*	*2010*	*2011*	*2011-2010*	*2011*	*2011*
HS 318	C. A. -2 (1 year)	15.0	1.8	-13.2	2.1	4.0
HS 290	S. I.- 1	15.0	0.4	-14.6	-2.2	-1.5
HS 319	S. I.- 2	12.4	6.4	-6.0	2.7	0.9
HS 312	S. I.- 2	1.6	1.8	0.2	6.2	5.1
HS 307	C. A.- 1	1.3	1.2	-0.1	3.2	-0.5
HS 306	S. I.- 1	1.3	2.0	0.7	4.4	-4.9
HS 302	S. I.- 2	0.0	-0.2	-0.2	2.2	1.6
HS 296	S. I.- 1	-0.5	-0.6	-0.1	-0.1	-1.8
HS 300	C. A.- 1	-0.6	1.2	1.8	0.6	3.9
HS 289	S. I.- 1	-2.2	0.0	2.2	0.6	0.6
HS 303	S. I.- 1	-2.8	-0.2	2.6	0.9	4.0
HS 288	S. I.- 1	-4.0	-2.7	1.3	1.3	-1.8
HS 291	S. I.- 1	-4.4	-0.1	4.3	3.2	-3.8
	Total	2.47	0.85	-1.62	1.93	0.45

Note: n=13. S. I. -1= School Improvement 1; S . I. -2= School Improvement 2; C. A. 1= Corrective Action 1; C. A. -2 (1 year)= Corrective Action 2 (1st year). 2011-2010= difference between AGI scores from 2011 to 2010.

Table 24

School Characteristics for the 2010 Making Progress Schools that Missed AYP in 2011: Reading

		PVAAS AGI			*PVAAS AGI*	
		Reading			*Science*	*Writing*
School ID	*2009 NCLB Status*	*2010*	*2011*	*2011-2010*	*2011*	*2011*
HS 322	S. I.- 2	15.0	0.7	-14.3	7.4	1.4
HS 317	S. I.- 1	15.0	2.1	-12.9	11.1	1.3
HS 308	C.A.2 (2nd year)	15.0	0.7	-14.3	-9.6	-2.5
HS 298	S. I.- 1	15.0	-4.9	-19.9	-0.1	-11.1
HS 321	S. I.- 1	7.5	6.0	-1.5	8.7	13.1
HS 315	S. I.- 2	6.5	5.9	-0.6	3.5	7.9
HS 320	C.A.2 (2nd year)	6.1	5.4	-0.7	0.4	1.0
HS 323	C. A. -2 (4+)	5.9	3.3	-2.6	-8.4	-1.5
HS 313	S. I.- 1	4.5	3.0	-1.5	-5.0	5.9
HS 309	S. I.- 1	3.6	-3.9	-7.5	-5.7	-2.6
HS 295	C.A. 1	3.4	-7.9	-11.3	1.9	-9.1
HS 310	C.A.2 (2nd year)	3.1	-0.1	-3.2	-3.0	0.3
HS 301	S. I.- 1	3.1	-2.1	-5.2	-2.9	-4.6
HS 314	S. I.- 1	2.3	2.6	0.3	2.2	-6.9
HS 324	C.A.2 (1st year)	2.0	4.4	2.4	1.1	1.9
HS 297	S. I.- 2	1.2	-0.6	-1.8	3.6	-2.6
HS 311	C.A. 1	-0.4	-3.5	-3.1	0.1	-0.3
HS 299	S. I.- 1	-0.7	-1.5	-0.8	-3.1	-2.0
HS 316	C.A. 1	-1.0	0.0	1.0	2.1	2.5
HS 287	C.A. 1	-1.2	3.2	4.4	-0.6	0.0
HS 292	S. I.- 2	-1.9	-2.4	-0.5	-5.2	-5.2
HS 286	S. I.- 1	-2.7	-1.1	1.6	-0.6	-1.8
HS 305	S. I.- 1	-3.1	-1.4	1.7	-1.6	-3.8
HS 304	S. I.- 2	-5.0	-6.5	-1.5	-3.7	-4.6
	Total	3.88	0.06	-3.83	-0.31	-0.97

Note: n= 24. The 2010 Making Progress schools that Missed AYP in 2011 would fall to one level below their 2009 NCLB Status (see Figure 1 for current NCLB status for these schools). S. I. -1= School Improvement 1; S. I. -2= School Improvement 2; C. A. 1= Corrective Action 1; C. A. -2 (1 year)= Corrective Action 2 (1st year); C. A. -2 (2 year)= Corrective Action 2 (2nd year); C. A. -2 (4+ year)= Corrective Action 2 (4 or more years). 2011-2010= difference between AGI scores from 2011 to 2010.

Table 25

2010 Making Progress Schools without PVAAS Results Reported in 2011: Reading

		PVAAS AGI Reading			*PVAAS AGI Science*	*PVAAS AGI Writing*
School ID	*2009 NCLB Status*	*2010*	*2011*	*2011-2010*	*2011*	*2011*
HS 294	-	15.0	-	-	-	-
HS 293	S. I.- 1	-3.5	-	-	-	-
		5.75	-	-	-	-

Note: n=2. HS 294 was closed in June 2011. HS 293 did not Make AYP during 2011. The 2010 Making Progress schools that Missed AYP in 2011 would fall to one level below their 2009 NCLB Status (see Figure 1 for current NCLB status for these schools). S. I. -1= School Improvement 1.

The line graph in Figure 9 depicts the 2010 and 2011 mean Reading AGI scores for the subgroup of 13 schools that Made AYP in 2011 (see Table 23) and the subgroup of 24 schools that Missed AYP in 2011 (see Table 24), from the group of 37 schools classified as Making Progress in 2010 that had data reported for both years.

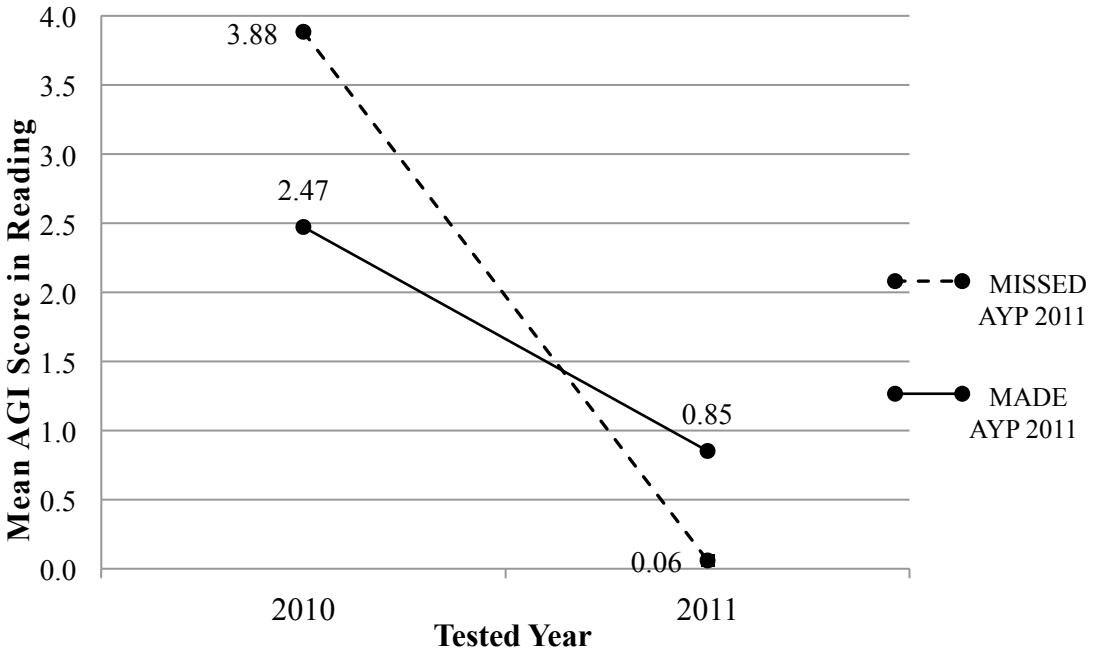

Figure 9. Mean AGI scores in Reading for 2010 and 2011, by schools that Made AYP and those that Missed AYP during the 2011 school year (restricted to schools classified as Making Progress in 2010)

As shown in Figure 9 both subgroups showed noticeable decline in mean AGI scores for Reading from 2010 to 2011, though the decline was more dramatic for the subgroup that Missed AYP in 2011.

A look at the seven schools with the highest score (+15.0) in Reading. Two general trends emerge from the results for the 2010 Making Progress group. Most of the schools failed to make AYP in 2011 (in other words, they reverted back to the Needing Improvement list). In like fashion, many of the schools showed significant decline in PVAAS Reading scores from 2010 to 2011. These effects are especially evident in the seven Making Progress schools that scored +15.0 in Reading in 2010. With regard to AYP status, only two of the seven schools Made AYP in 2011; four schools Missed AYP, and one school closed. From the data in Table 23 and Table 24, one can determine that the average reading score for the six remaining schools (those still in existence in 2011) dropped from +15.0 to +0.1, a decline of 14.9 points. It seems that having a high PVAAS score one year provides little protection against failure in the future.

Summary of Findings

PVVAS AGI scores for 426 Pennsylvania high schools were analyzed to advance three research hypotheses for the content areas of Mathematics and Reading. The following were the tested hypotheses and statements of findings:

Hypothesis One: There will be identifiable and statistically significant differences in high schools' PVAAS Average Growth Indices (AGI) in Mathematics and Reading as a function of School Improvement Classification (SIC). *Findings for Hypothesis One:* One-way ANOVAs reported significant differences of mean AGI scores between levels of the independent variable for both content areas, Mathematics and Reading.

Hypothesis Two: Schools identified as Making Progress in 2010 will have higher PVAAS Average Growth Indices (AGI) in Mathematics and Reading than schools identified as Needing Improvement. *Findings for Hypothesis Two:* Tukey post hoc analyses found four significant relationships in Mathematics; three of the four relationships involved the Making Progress group significantly outperforming other groups. Nine significant relationships were revealed in Reading; six of the nine involved the Making Progress group outperforming other groups.

While not specifically germane to Hypothesis 2, there is an additional finding worthy of note. The group of schools that had been under NCLB scrutiny and sanction the longest time (SIC B8, never Made AYP in 7 years) had the lowest PVAAS AGI means in both Mathematics and Reading.

Hypothesis Three: Schools Making Progress in 2010 will continue to demonstrate similar PVAAS gains in Mathematics and Reading during the 2011 testing year. *Findings for Hypothesis Three:* Paired sample t-tests revealed that PVAAS AGI scores for the 2010 Making Progress group dropped significantly between 2010 and 2011 in both Mathematics and Reading; these findings contradicted Hypothesis 3.

The data for the 2010 Making Progress group were further analyzed by AYP outcome in 2011. In Mathematics, some schools sustained their good performance from 2010 into 2011; this was more characteristic of schools that Made AYP in 2011. Other schools showed a considerable decline in Mathematics PVAAS scores from 2010 to 2011; this was more characteristic of schools that Missed AYP in 2011. The story is somewhat different for the Reading results. Both subgroups, those that Made AYP and

those Missed AYP in 2011, showed considerable decline in Reading PVAAS AGI scores from 2010 to 2011 (though the decline was less steep for the Made AYP subgroup).

A final observation warrants mention. Data from the 2010 Making Progress group suggest that higher is not always better (referring to PVAAS AGI scores). The subgroup of schools that Missed AYP in 2011 had higher PVAAS scores in 2010 (in both Mathematics and Reading) than the subgroup that Made AYP in 2011. Furthermore, the schools with the highest obtained score in Reading (+15.0) in 2010 dropped, on average, an astounding 14.9 points in 2011.

The next chapter will discuss these findings and their implications for educational policy and research.

CHAPTER 5

Discussion and Conclusion

No Child Left Behind (NCLB) ushered in transformations to the Elementary and Secondary Education Act (ESEA), most notably through unprecedented legislated accountability for all public schools (Tosh & Edwards, 2009). Since the law's inception, schools have worked to increase student outcomes on high-stakes tests to avoid public embarrassment, sanctions, and the threat of state take-over. Under this system of mandated external accountability, high schools have unique challenges that make improvement difficult to achieve and sustain over time (Duke & Jacobson, 2011). However, nearly ten years after the inception of NCLB, an issue that remains unsettled is whether NCLB can accurately identify schools that need improvement and then support those schools at high enough levels to increase performance over time.

The purpose of this study was to use value-added methodologies to determine if external accountability, as defined by NCLB, is improving student achievement on 11th grade Mathematics and Reading assessments in Pennsylvania high schools. Scaffolded research questions along with research hypotheses were used to first determine whether the Pennsylvania Value Added Assessment System (PVAAS), as a statistical instrument, successfully quantified the improvement of schools across the School Improvement Classifications (SIC). Second, to the extent that PVAAS quantified the improvement of schools, the group of schools identified as Making Progress should demonstrate higher gains than schools Needing Improvement to infer a positive relationship between external accountability and school performance. The final step was to determine whether the schools identified in 2010 as Making Progress continued to improve in 2011. Statistical

tests were carried out for each step (defined by research hypotheses) for the content areas of Mathematics and Reading. This chapter will summarize the findings of the data analyses, provide a discussion of those findings, and give recommendations for practice and further research.

Summary and Discussion of Key Findings

The present study examines the influence of NCLB's mandates over school improvement by testing a school's ability to improve both its standing within the existing system of external accountability, as defined by School Improvement Classifications (SIC), and the performance of students in Mathematic and Reading, as defined by PVAAS scores. Essentially, deducing NCLB's positive influence on school improvement would require schools to improve both their NCLB status and student performance (PVAAS AGI scores). This study will contribute to the existing literature and ongoing discussion regarding the influence of mandated accountability over school improvement. The following are the summarized findings from this study:

Hypothesis One

The first hypothesis tested whether PVAAS' Average Growth Index (AGI) captured the improvement of schools that were identified by NCLB's methodology as "needing improvement." In other words, if the PVAAS AGI scores differed significantly across School Improvement Classifications, it might imply that the sanctions and school improvement processes had a role in both improving the NCLB status of schools as well as increasing student performance on high-stakes tests. The one-way analysis of variance (ANOVA) found significant differences for both content areas (Mathematics and

Reading). Significant findings for the first hypothesis were needed to advance the statistical analyses to next step of this study.

Hypothesis Two

The next step in this study was to determine if the Making Progress group outperformed other groups. Schools comprising the Making Progress group have two notable characteristics that are central to this study. First, Making Progress schools *were* identified the previous year as Needing Improvement using NCLB's methodology and subsequently endured the sanctioning and school improvement process established by NCLB and the state of Pennsylvania. To infer that NCLB had an influence on school performance, it would be necessary for "improving" schools to experience the sanctions and school improvement process first-hand. The second characteristic that sets the Making Progress schools apart from the others is that these schools have improved something to the extent that they are not currently labeled as Needing Improvement within the existing system of accountability. Again, to assume a relationship between NCLB and school performance, schools would need to demonstrate advancement within the existing system of external accountability.

With regard to the second hypothesis, there were four significant findings for Mathematics. Two of these four findings involved the Making Progress group (SIC B3) significantly outperforming the following two groups needing improvement: 1) the group that "Missed AYP for the first year" (SIC B4) and 2) the group that "never Made AYP" (SIC B8). These results were in-line with this hypothesis and could mean that the effects of NCLB influenced not only NCLB status for these schools but also their PVAAS AGI scores.

In addition to out-performing two groups Needing Improvement, the Making Progress group also significantly outperformed schools that "Always Made AYP." This result was not directly aligned to Hypothesis 2 but could imply that NCLB's accountability had a part to play here as well. For example, NCLB as a school improvement platform assumes that consequences, in the form of incentives and sanctions, will motivate schools to improve (Finnigan & Gross, 2007). In this manner, schools that "Always Made AYP" might simply be less motivated to improve than the Making Progress group.

The fourth significant finding for Mathematics, involved the group that "Made AYP for 2009 and 2010 (but previously Missed AYP)" (SIC B2) outperforming the group that "Missed AYP (for more than 7-years)" (SIC B8). Again, this finding could also support the second hypothesis because schools comprising the group "SIC B2" would have been sanctioned or participated in a mandated school improvement process in the past. Essentially, the "SIC B2" group would include all the Making Progress schools that continued to Make AYP in subsequent years.

In the content area of Reading, the Making Progress group significantly outperformed all other groups with the exception of the group of schools that "Missed AYP (for 2 or 3 years)." In addition, the group that "Missed AYP (for more than 7-years)" performed significantly worse than all the groups that made AYP (SIC B1, B2, and B3) as well as the group of schools that "Missed AYP (4 or 5 years)" (SIC B6). In total, there were nine significant findings for Reading and six of the nine involved the Making Progress group outperforming other groups. The findings for this second hypothesis seemed to coincide with the notion that external accountability influenced

both the status of schools within the system of accountability as well as the outcomes of students on high-stakes tests, as measured by PVAAS.

Hypothesis Three

The last step in this study sought to determine whether the 2010 Making Progress group continued to maintain their NCLB accountability status as well as make gains represented by PVAAS in subsequent years of testing (2011). Elmore (2007) offers a simple definition for school improvement as "increases in quality and performance over time" (p. 221). Reeves (2004) describes accountability as a cycle of *continuous* improvement. Similarly, drawing on a decade of school improvement research, Anderson and Kumari (2009) claim *continuous* school improvement requires more than policy mandates, but instead schools will need to become learning organizations that constantly analyze both student results and teacher practices within an aligned system of accountability.

> …even in schools that experience the greatest pressure to improve, we do not have a good idea of how many actually undertake and demonstrate continuous improvement on government prescribed performance measures over time, and what this means in practice. (Anderson & Kumari, 2009, p. 282)

Goodwin (2011) characterizes continually high performing schools as simply, "more reliable" (p. 136). According to Goodwin, "…schools do not need to be flashy or up on the latest trends to be effective. They just need to ensure that students receive the same high-quality learning experiences in every classroom" (p. 136). Hypothesis 3 intended to separate the more reliable schools, those schools that continued to improve in 2011, from those that had glimpses of success but were unable to sustain the success over time.

The positive results from the first two hypotheses gave prospect that external accountability might account for *some* of the improvement demonstrated by high schools in Pennsylvania. Optimism was attributed to the 2010 Making Progress group reporting significantly higher AGI scores than schools Needing Improvement for both Mathematics and Reading. However, the third hypothesis employed paired-sample t-tests that revealed significant declines for both Mathematics and Reading from 2010 to 2011. Thus, Hypothesis 3 was rejected, along with notion that NCLB influenced the sustained improvement of Pennsylvania high schools.

The data for the 2010 Making Progress group were further analyzed by AYP outcome in 2011. The original 2010 Making Progress group had 39 schools. However, two of these schools did not report PVAAS AGI scores in 2011 and were removed from all analyses associated with Hypothesis 3. Of the 37 schools that did report scores in 2011, 13 schools made AYP and 24 schools missed AYP. In Mathematics, the AGI scores for the group of schools that Made AYP in 2011 increased slightly (from +1.52 in 2010 to +1.62 in 2011). However, the scores for the group of 24 schools that Missed AYP in 2011 dropped 2.06 points (from +3.80 in 2010 to +1.74 in 2011). Since the group that Made AYP in 2011 was able to sustain progress in Mathematics from the 2010 testing year to 2011, it could be interpreted that the system of external accountability had its intended effect on these 13 schools.

In Reading, neither group sustained improvement into the 2011-testing year. The group of 13 schools that Made AYP in 2011 declined in Reading by 1.62 points (from +2.47 in 2010 to +0.85 in 2011). The group of 24 schools that Missed AYP in 2011 fell 3.83 points (from +3.88 in 2010 to +0.06 in 2011). Unlike Mathematics however, where

the overall decline in scores on the 2011 assessment could be *somewhat* rationalized, the Reading PVAAS AGI scores could not. It was clear that the schools that scored the highest in Reading in 2010 seemed to drop the furthest on the 2011 assessment.

The seven schools with the highest score in Reading (+15.0) during 2010 were most perplexing. Remarkably, all seven of these schools landed in the 2010 Making Progress group. Of the seven, only six reported PVAAS scores in 2011. For these six schools, reading scores dropped an average of 14.9 points, going from +15.0 in 2010 to +0.1 in 2011. Essentially, the high scores in 2010 for these schools were entirely erased during the subsequent testing year. The issues related to how and why the highest performing schools in 2010 would decline in such an extreme manner in 2011 are examined further in the next section.

Discussion of Key Issues

This discussion addresses two key issues from this study. The first section involves the use of PVAAS as a measurement of school improvement. Specifically, the discussion in this area will address whether PVAAS, as a statistical measurement of school performance, can effectively quantify the improvement of schools. The second section addresses the utility of mandated accountability itself. Drawing on the theory of accountability, the discussion in this area will address NCLB's usefulness as a mechanism for school improvement.

Value Added Assessments

At the time of this study, NCLB used "mean proficiency scores" on high-stakes tests in Mathematics and Reading to determine the NCLB status of a school. Under this system, schools that meet arbitrary proficiency targets are considered successful while

schools that fall short are labeled as "needing improvement." However, researchers report that this "traditional" approach to identifying schools for improvement is lacking and fails to accurately measure the improvement of schools over time or apart from covariates that influence achievement (Davies, 2008; Forte, 2010; Harris, 2010; Hershberg, Simon, & Lea-Kruger, 2004; Linn, 2008; Meyer, 2000; Raudenbush, 2004).

According to Reeves (2005), "One of the most important developments in educational accountability has been the 'value-added' methodology developed by Professor William Sanders and his colleagues at the University of Tennessee" (Reeves, 2005, p.16). The Pennsylvania Value-Added Assessment System (PVAAS) employs a version of the system developed by Sanders. The key difference between PVAAS and traditional methodologies is that PVAAS uses a longitudinal model to calculate the "growth" or gain of individual students, which are then aggregated at the level of the individual schools. The traditional approach, on the other hand, is based on "cross-sectional" comparisons using different cohorts of students.

Currently, Pennsylvania uses the value-added methodologies of PVAAS as an alternate means of achieving Adequate Yearly Progress (AYP) in grades four through eight by projecting students' scores on future assessments and then using those projections as indicators of proficiency for the current year. Since high schools are unable to project scores beyond the 11th grade assessment, PVAAS's predictive methodologies cannot be used as a means to achieve AYP in Pennsylvania high schools. The PVAAS comparison data (regarding the 2010 assessment) for all Pennsylvania high schools was made public, thus providing the opportunity to examine the effect of

mandated accountability on Pennsylvania high schools with a presumably more reliable indicator of school improvement.

The findings from this study might raise questions from members of the research community regarding the reliability of PVAAS AGI scores as a means to determine school performance. One unsettling issue arising from the analyses conducted for this study involves the extreme range of AGI scores for the distribution as a whole and within each of the School Improvement Classifications.

As shown in Table 26, PVAAS AGI scores in Mathematics ranged from -10.6 to +20.0 for the full data set. Table 26 also shows the minimum and maximum score for each School Improvement Classification.

Table 26

Range of PVAAS Scores in Mathematics

Group/SIC Code	Minimum Score	Maximum Score
All Data	-10.6	20.0
B1	-9.2	15.4
B2	-10.6	16.4
B3	-7.1	12.0
B4	-7.0	20.0
B5	-7.1	7.1
B6	-5.8	11.6
B7	-7.3	10.4
B8	-9.1	9.7

Table 27 shows the minimum and maximum Reading scores for the full data set and within each School Improvement Classification.

Table 27

Range of PVAAS Scores in Reading

Group/SIC Code	Minimum Score	Maximum Score
All Data	-10.5	15.0
B1	-7.4	7.1
B2	-8.0	9.4
B3	-5.0	15.0
B4	-8.6	9.4
B5	-9.4	6.0
B6	-4.0	7.0
B7	-4.4	5.6
B8	-10.5	7.1

To put these scores in perspective, Table 28 provides the interpretation guidelines for PVAAS scores that are found on the PVAAS public website at https://pvaas.sas.com/evaas/signin.jsf. According to these guidelines a +2.0 and higher shows "significant evidence" that the school exceeded the standard for growth. Further, implicit within the interpretation of these scores is the perception of "relative standing" or ranking, where the higher the AGI score the better the performance. Therefore, how does one interpret an exceedingly high AGI score of +15.0, if a +2.0 is exceedingly high? Similarly, what can be made of a -10.0, when an AGI score of -2.0 is interpreted as exceedingly low?

Table 28

Interpretation of 2011 PVAAS Scores

2011 AGI Cut-Point	Interpretation
2.0 and Higher	Significant evidence that the school exceeded the standard for PA Academic Growth
1.00 to 1.99	Moderate evidence that the school exceeded the standard for PA Academic Growth
-0.99 to 0.99	Evidence that the school met the standard for PA Academic Growth
-1.00 to -1.99	Moderate evidence that the school did not meet the standard for PA Academic Growth
-2.0 and Below	Significant evidence that the school did not meet the standard for PA Academic Growth

Note: Adapted from the 2011 PVAAS public Reporting Website at: https://pvaas.sas.com/evaas/welcome.jsf
AGI= Average Growth Index

The point can be taken further. In Mathematics, 139 of the 426 schools (or 32.6%) had PVAAS AGI scores +2.0 and higher; 137 schools (or 32.2%) had PVAAS AGI scores -2.0 and below. So by this estimate, 64.8% of schools are exceptional, one way or the other.

In Reading, 114 of the 426 schools (or 26.8%) had PVAAS AGI scores +2.0 and higher; 108 schools (or 25.4%) had PVAAS AGI scores -2.0 and below. Here, 52.2% of schools are judged to be exceptional.

Perhaps most puzzling are the schools whose Reading scores dropped, on average, from +15.0 to +0.1 in the span of a single year. Apart from assessment misconduct (cheating), which will only be addressed as a recommendation for further research, the erratic PVAAS Reading scores, reported in this study, could be the result of

incomplete data-sets beset by high student mobility, errors in data processing, or incomplete record-keeping at the state or local level (Amrein-Beardsley, 2008).

Amrein-Beardsley (2008) insisted that more criterion-related validity tests are required before using PVAAS's methodology for teacher evaluations. At the time of this study, the methodology espoused by PVAAS is already contributing to teacher evaluations in other states. For example, in May 2011, the Los Angeles Times obtained access to school district records and posted the names of 11,500 teachers along with their value-added scores (Song & Feltch, 2011). Clearly, stakes have never been higher; additional validity tests, using independent measures of teacher quality, are needed before inferring causal relationships between PVAAS gains and teacher effectiveness.

As with teacher evaluations, more criterion-related validation studies are also needed at the school-level. For example, comparisons between independent credentialing agencies (i.e., Middle States Standards for Accreditation for Schools and National Blue Ribbon Schools Program) should be conducted as a baseline for validity. Moreover, comparisons to other, lower-stakes tests such as NAEP, Advanced Placement, SATs and ACTs are needed to establish validity of findings.

Criterion-referenced validity assessment. One way to assess the validity of PVAAS as a measuring instrument is to compute the correlation between PVAAS scores and some criterion measure of success. Table 29 reports the relationship between PVAAS scores (on both high-stakes and low-stakes tests) and the criterion of Making vs. Missing AYP in 2011. All correlations are based on the 37 schools classified as Making Progress in 2010 that had scores reported for both 2010 and 2011 (the data appear in Table 14, Table 15, and Table 16).

Table 29

Correlations between PVAAS scores (from 2010 and 2011) and AYP status (2011) for 37 Schools Classified as Making Progress in 2010

	Criterion-Referenced Validity Correlations					
	Predictive Validity 2010 to 2011		Concurrent Validity 2011 to 2011			
Point-biserial Correlations	Math	Reading	Math	Reading	Science	Writing
(Pearson model) AYP status 2011	-.243	-.109	-.014	.116	.250	.149
(Spearman variation) AYP status 2011	-.257	-.191	.000	.093	.326 *	.202

Note. * Correlation is significant at the .05 level (2-tailed). AYP= Adequate Yearly Progress

The criterion variable is AYP status 2011 (1 = Made AYP, 0 = Missed AYP). The indicator variables are the PVAAS Mathematics and Reading scores for 2010 (predictive validity) and the PVAAS Mathematics, Reading, Science, and Writing scores for 2011 (concurrent validity). For the Pearson variation of the point-biserial correlation, the PVAAS score is continuous and can take on a wide range of values; these are the scores we have used and analyzed throughout this study. For the Spearman variation of the point-biserial correlation, the original PVAAS scores are converted to rank orders. This mitigates the distortion arising from outlier values.

Mathematics and Reading are called "high-stakes" because performance on these tests determines whether a school Makes or Misses AYP in a given year. Science and Writing are considered "low stakes" tests because they currently play no role in determining AYP status.

Only one of the correlations reached the level of statistical significance. The following descriptive analysis is presented with the aim of hypothesis generation rather

than hypothesis testing. The goal is to provide promising avenues for further research in this area.

Predictive validity. The correlation between 2010 PVAAS scores (Mathematics or Reading) and 2011 AYP status was negative. Schools with higher PVAAS scores in 2010 in these content areas were somewhat less likely to Make AYP in 2011 than schools with lower PVAAS scores.

Concurrent validity. The correlation between 2011 PVAAS scores for "high-stakes" Mathematics and Reading tests and 2011 AYP status was essentially zero. PVAAS Mathematics and Reading scores in 2011 had no bearing on AYP outcome in 2011.

The correlation between 2011 PVAAS scores for "low-stakes" Science and Writing tests and 2011 AYP status was positive. Schools with higher PVAAS scores on the "low-stakes" tests were somewhat more likely to Make AYP in 2011 than schools with lower PVAAS scores.

Implication of findings. According to these findings, tests with lower-stakes (Writing and Science) are better predictors of AYP status than high-stakes tests (Mathematics and Reading). These findings directly challenge the theory of test-based accountability, which is predicated on the assumption that increasing "stakes" (sanctions, rewards, etc.) yield increases in both student performance (PVAAS scores) and the school's standing in the existing system of accountability (AYP Status). These preliminary results might support the findings from Amrein and Berlinger (2002), reported earlier in this study, suggesting that the strength of an accountability policy could actually have a negative impact on student performance. Clearly, more information

is needed before any inference can be made regarding the validity of PVAAS. However, these findings do point to a generative strand of further research in this area.

Section summary. While many researchers believe that PVAAS is an improvement over the traditional method that uses mean proficiency scores and arbitrary targets to determine school improvement, validation research for value-added methodologies is limited and largely inadequate (Amrein-Beardsley, 2008; McCaffrey, Lockwood, Koretz, & Hamilton, 2003). More validation tests, such as the example reported in this section, are needed before PVAAS can legitimately be used as the sole or primary basis for determining which schools (or teachers) should receive reward or sanction.

Did NCLB "put the cart before the horse" by establishing the mandate for school improvement without a stable methodology to evaluate school performance? Based on the existing research in this area, it seems likely. The bigger, and yet unanswered, question is whether new value-added methodologies improve the traditional method enough to support the national mandate for accountability. The next section will address the utility of external accountability to improve the performance of public schools.

External Accountability to Improve Schools

The core of this study examines whether a national system of accountability can improve the performance of individual schools. The previous section questioned the extent that PVAAS, as a statistical instrument, was stable and precise enough to capture the improvement of schools already in a system of external accountability. Specifically, attention was drawn to the wide-ranging (erratic) PVAAS scores within the entire distribution as well as within each classification of school improvement (see Table 26

and Table 27 for examples). Also, addressed was the "perplexing" situation regarding the plummeting Reading scores from the highest scoring group of schools in 2010 (declining from +15.0 in 2010 to +0.1 in 2011). However, guided by the seminal research conducted by Elmore (2007), this second section pursues an alternate explanation regarding the plummeting test scores from the highest-ranking schools on the 2010 Reading assessment. Specifically, this section submits that the existence of high-stakes, external accountability, itself, might provide an explanation why all of the highest performing schools on the 2010 Reading assessment landed in that year's Making Progress group, and why only two of those schools Made AYP in 2011.

According to Elmore (2007), most schools follow a common pathway toward reform that generally begins with schools recognizing problems through an examination of evidence around student performance. After realizing that a problem exists, schools typically establish a goal as well as a strategy for attaining the goal. Elmore's research discovered that schools often see a modest "bounce" in performance (p. 249) during this phase. For schools with low levels of organizational capacity the bounce is likely due to specific low-level strategies, such as teaching or re-teaching specific skills to all students or targeting specific students prior to the high-stakes assessment. Elmore considers these initial, low-level strategies the "some teaching vs. no teaching" phase of school reform (p. 250).

Regarding the Reading results from the present study, the high scores obtained by the seven highest performing schools in 2010 might be the direct result of specific low-level changes implemented during the "some teaching vs. no teaching" phase. After all, Elmore (2007) emphasized that initial, low-level changes "always turn out to have very

short-term, very disappointing effects" (p. 250). In this sense, the very disappointing performance of the 2010 Making Progress group on the 2011 Reading assessment gives credence to this theory. More importantly, Elmore continues that the initial stage of reform is the "critical moment" (p. 250) where schools either develop mechanisms for internal accountability by establishing common expectations and building collective capacity or simply continue to grapple for the "low hanging fruit."

During the planning stages of this study, the author of this dissertation thought that the more time a school spent within the system of school improvement, the more growth those school might experience through PVAAS. However, Elmore (2007) explains that the effect of external accountability systems, that which we are currently exploring, is mediated by the internal accountability mechanisms within the school. "In other words, how a school responds to external accountability systems is largely determined not by the details of the accountability system but by the degree of alignment between the school's internal accountability mechanisms and the requirements of external systems" (p. 143). In this sense, the effect of the NCLB mandate on the improved quality of a school should, in some degree, be observed through the performance of the school within the existing system (AYP status), as well as on student performance (PVAAS AGI scores).

The disappointing performance of the 2010 Making Progress group on the 2011 assessments may ultimately be the result of a "misalignment" between the school's internal accountability mechanisms and external systems rather than methodological issues associated with PVAAS. The fact that only 13 schools from the initial cohort were able to meet the requirements of AYP in 2011 might suggest either higher levels or more

aligned mechanisms of internal accountability for these schools. Contrariwise, the significant decreases demonstrated by the other schools on the 2011 Reading assessment might imply that these schools picked all the "low-hanging fruit" and were essentially "tapped out," due to an inherent lack of capacity. If this is the case, Newmann, King and Rigdon (1997) describe a bleak situation where external accountability actually stifles the capacity-building efforts for schools with already low capacity. Elmore (2007) confirms that developing the capacity of schools is challenging for reforming schools and is a likely reason that many schools are unable to sustain improvement over time. According to Elmore and supported by other researchers in this field, the only salvation for low performing, low capacity schools in a system of external accountability is to seek technical assistance from outside agencies (Elmore, 2007; Newman, King, & Rigdon, 2007). In short, failing schools need help from the outside to improve, after all "by definition, people in the school don't know what to do, or they would have done it already" (Elmore, 2007, p. 250).

Section summary. The performance of the 2010 Making Progress group may ultimately have everything to do with NCLB's external accountability, both through increases in 2010 and subsequent decreases in 2011. Even though accountability policies might provide the impetus for low-level changes in schools during the short-term, "test-based accountability without substantial investments in capacity – internal accountability and instructional improvement in schools – is unlikely to elicit better performance from low-performing students and schools" (Elmore, 2007, p. 207). Based on this assumption, external accountability without the accompanying technical support for low-performing schools may not only cause but perpetuate erratic patterns of performance characterized

by the schools in this study that scored exceedingly well in 2010 but dropped notably the next year.

Recommendations for Practice

The findings from this study point to recommendations that will support teachers, school principals, district administrators, policy makers, and researchers in education. These recommendations should both improve how schools respond to accountability and what public officials can do to enhance the existing system of accountability in education.

According to Harris (2010), "value-added measures of school performance are clearly superior to attainment measures in order to maximize fairness" (p. 68). However, school leaders should use caution when interpreting the findings from value-added assessments. Without more validation studies, inferences regarding school or teacher performance should not be made solely on the value-added methods discussed in this study.

Policy makers should redirect reform efforts toward creating an infrastructure that can support schools at high enough levels with the goal of sustaining performance over time. Support in this area will likely include building the capacity of teachers around shared norms, values, and expectations. According to Elmore (2007), adequately providing the tools and expertise for schools to attain proficiency is developed through the concept of "reciprocity." District and state leaders, should evoke the "principle of reciprocity" (Elmore, 2007, p. 251) and demand more high-quality, technical support from federal agencies in return for increases in performance. Without reciprocity, failing schools will be unable to meet the demands of external accountability (Elmore, 2007).

Policy makers involved in the reauthorization of the Elementary and Secondary Education Act (ESEA) should consider alternate models to evaluate schools beyond high-stakes testing. According to Forte (2010) significant changes to the existing accountability model are not likely. Therefore, including performance-based assessments and focusing on teacher quality and capacity building could be improvements more easily made within the existing system of accountability. Also, developing the capacity of failing schools is going to necessitate "a lot of feet on the ground" (Elmore, 2007, p. 255). Therefore, policy makers should consider different approaches to disseminating federal dollars to needy schools. For example, rather than distributing money through federal block grants that necessitate a bureaucratic system to manage and track funds, dollars should go directly to the schools and districts that need it.

Recommendations for Further Research

Although this study provides new information, further research is required to both corroborate and advance these reported findings. This section will present recommendations for further research identified through this study.

The PVAAS methodology requires additional independent validity tests regarding its legitimate and ethical use as a measurement of school performance. Validity tests should include criterion-related analyses using other measures of school performance as well as construct-related analyses to ascertain whether the PVAAS methodologies accurately assess student learning and to what extent (Amrein-Beardsley, 2008).

This study addressed the implications of mandated accountability on Pennsylvania high schools that only administer the eleventh grade PSSAs. Similar, quantitative studies using state-wide or regional samples of middle schools (grade 6-8)

are needed to determine the effects of external accountability on schools that administer high-stakes tests to all students, each year.

The Pennsylvania System of School Assessment (PSSA) uses forensic tests to investigate potential test misconduct (cheating). As policy-makers advance the application of value-added models to make inferences regarding teachers, instructional practices, and curriculum, more research is needed to explore the findings from these forensic reports in Pennsylvania. Specifically, comparisons should be made between the forensic findings from high-stakes and low-stakes assessments to determine the extent high-stakes accountability policies contribute to assessment misconduct.

Findings from this quantitative study suggest that low-stakes tests might serve as better concurrent predictors of AYP status than high-stakes tests. Mixed-method studies are now needed to both compare schools' performance on high-stakes vs. low-stakes tests and then investigate the extent schools prepared students for these tests using particular strategies. Specifically, this area of research will seek answers to the following: 1) Do schools engage in "test-prep" or "target specific sub-groups" for either high-stakes or low-stakes tests? 2) What strategies do successful schools practice? 3) What strategies do unsuccessful schools practice? 4) How much instruction time is compromised for such practices? 5) To what extent does employing (or avoiding) specific strategies influence schools' AYP status? 6) To what extent does employing (or avoiding) specific strategies influence student outcomes?

According to Troen and Boles (2012), "Without the principal's vision all that occurs in schools will be transitory. Nothing – and we mean nothing- can become long-lasting in a school without the initial and continued support of the principal" (p.27).

Steady and coherent principal leadership is critical to reforming schools. Quantitative research studies, examining the influence of principal-turnover (especially in urban schools that are improving), are needed to determine the mediating effects of *consistent* leadership on continuously improving (urban) schools.

Schmoker (2006) writes that the best School Improvement Plans (SIP) are straightforward procedures that provide opportunities for teachers to get together, analyze student achievement data, and set targets based on the data. All Pennsylvania schools that do not make AYP are required to submit SIPs. Additional, quantitative and qualitative studies regarding what makes a SIP effective would bring tremendous value to all schools going through the improvement process.

Tools that measure collective capacity within an organization are already employed in education (Newmann, King, & Rigdon, 1997). Researchers should further use these tools to identify schools with high-levels of collective capacity and determine how those schools developed their capacity over time. Finally, the Pennsylvania Department of Education should develop partnerships with higher education to examine the vast stockpiles of data archived within the public domain. Research grants should be extended to several state-based research facilities with the sole purpose of generating high-quality quantitative analyses of the existing data, such as PVAAS and student performance patterns on high-stakes tests. A partnership at this level would be dedicated to a common purpose, improving the sustained quality of schools.

Conclusion

The findings from this study are mixed. Some findings show that changes in NCLB status go hand-in-hand with changes in PVAAS AGI scores. Other findings show

that PVAAS scores are erratic and unreliable. Even if Value-Added Models (VAM) are considered an upgrade over traditional methods for computing school improvement, more validity testing is needed before making inferences regarding school quality (or teacher quality), based solely on this methodology.

Some findings suggest that the current system of external accountability in education is working, while other findings suggest the system is largely ineffective and quite possibly even hampering improvement, especially in those schools with already low-levels of student proficiency. Schools that never Made AYP (in the past 7 years) experienced the most sanctions under NCLB, but had little to show for it; they had the lowest PVAAS scores of all categories. Even schools that made their way off the Needing Improvement list were likely to fall back on. Two-thirds of the schools in the 2010 Making Progress group failed to Make AYP in 2011, and their mean PVAAS scores dropped in turn.

The significance of internal accountability and organizational capacity within schools (as opposed to external accountability mandated from afar) may ultimately provide a means to reconcile the seemingly disparate findings from this study, while pointing to key areas of improvement that apply to all schools.

References

Abelmann, C., Elmore, R., Even, J., Kenyon, S., & Marshall, J. (1999). When accountability knocks will anyone answer? Retrieved on April 23, 2011 from Consortium for Policy Research in Education website: http://www.cpre.org/images/stories /cpre_pdfs/rr42.pdf

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives,* 10(18). Retrieved 9/11/2011from: http://epaa.asu.edu/epaa/v10n18/

Amrein-Beardsley, A. (2008). Methodological concerns about education value-added assessment system. *Educational Researcher, 37*(2), 65-75. Doi:10.3102/0013189X08316420

Amrein-Beardsley, A. (2009). Value-added tests: Buyer be aware. *Educational Leadership, 67*(3), 38-42.

Anderson, S., & and Kumari, R. (2009). Continuous improvement in schools: Understanding the practice. *International Journal of Educational Development, 29,* 281-292.

Ary, D., Jacobs, L.C., & Sorensen, C. K. (2010). *Introduction to research in education*. Belmont, CA: Wadsworth Cengage Learning.

Balfanz, R., Legters, N., West, T. C., & Weber, L. M. (2007). Are NCLB's measures, incentives, and improvement strategies the right ones for the nation's low-performing schools? *American Education Research Journal, 44*, 559-593. doi:10.3102/0002831207306768

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics, 29*, 37-65.

Brodsky, A. (2008). *Accountability reform and student achievement in Colorado public schools*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Publication No AAT 3337065)

Bryk, A. S. (2010). Organizing schools for improvement. *Phi Delta Kappan*, *91*, 23-30.

Carnoy, M. (2005). Have state accountability and high-stakes tests influenced student progression rates in high school? *Educational Measurement: Issues & Practice, 24*(4), 19-31. doi:10.1111/j.1745-3992.2005.00020.x

Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Education Evaluation and Policy Analysis, 24*, 305-331.

Carpenter, W. A. (2000). Ten years of silver bullets: Dissenting thoughts on education reform. *Phi Delta Kappan*, *81*, 383-389.

Chatterji, M. (2002). Models and methods for examining standards-based reforms and accountability initiatives: Have the tools of inquiry answered pressing questions on improving schools? *Review of Educational Research, 72*, 345-386. doi: 10.3102/00346543072003345

Cobb, C. D., & Rallis, S. F. (2008). District responses to NCLB: Where is the justice?. *Leadership and Policy in Schools*, *7*, 178-201. doi:10.1080/15700760701748378

Cuban, L. (2004). Looking through the review mirror at school accountability. In K. A. Sirotnik (Ed.), *Holding accountability accountable: What ought to matter in public education.* New York, NY: Teachers College Press.

Cuban, L. (2007). Hugging the middle: Teaching in an era of testing and accountability, 1980-2005. *Education Policy Analysis Archives*, *15*(1), 1-27. Retrieved from the Education Policy Analysis Archives website: http://epaa.asu.edu/epaa/v15n1/.

Darling-Hammond, L. (1989). Teacher professionalism and accountability. *The Education Digest, 55*(1), 15-20.

Darling-Hammond, L. (2004). Standards, accountability, and school reform. *The Teacher College Record, 106*, 1047-1085.

Davies, P., Coates, G., Hammersley-Fletcher, L., & Mangan, J. (2005). When 'becoming a 50% school' is success enough: A principal-agent analysis of subject leaders' target setting. *School Leadership and Management, 25*, 493-511.

Davies, R. S. (2008). AYP accountability and assessment theory conflicts. *Mid-Western Educational Researcher, 21*, 2-8.

Duke, D. L., & Jacobson, M. (2011). Tackling the toughest turnaround – low-performing high schools. *Phi Delta Kappan, 92*(8), 34-38.

Elmore, R. F. (2005). Accountable leadership. *Educational Forum, 69*, 134-142.

Elmore, R. F. (2007). *School reform from the inside out: Policy, practice, and performance.* Cambridge, MA: Harvard Education Press.

Elmore, R. F., & Fuhrman, S. H. (2001). Holding schools accountable: Is it working?. *Phi Delta Kappan, 83*(1), 67.

Fernandez, K. E. (2009). Evaluating school improvement plans and their effect on academic performance. *Educational Policy, 25*, 338-367.

Ferris, J. M. (1992). School-based decision making: A principal-agent perspective. *Educational Evaluation and Policy Analysis, 14*, 333-346.

Finnigan, K. S., & Gross, B. (2007). Do accountability policy sanctions influence teacher motivation? Lessons from Chicago's low-performing schools. *American Education Research Journal, 44*, 594-629. doi:10.3102/0002831207306767

Forte, E. (2010). Examining the assumptions underlying the NCLB federal accountability policy on school improvement. *Educational Psychologist, 4,* 76-88.

Fraenkel, J. R., & Wallen, N. E. (2009). *How to design and evaluate research in education*. New York, NY: McGraw-Hill.

Fried, L. (2009). Education, language and professionalism: Issues in the professional development of early years practitioners in Germany. *Early Years, 29*(1), 33-44.

Fuhrman, S. H. (1999). *The new accountability.* Retrieved on April 23, 2011 from Consortium for Policy Research in Education website: http://www.cpre.org/images /stories/cpre_pdfs/rb27.pdf

Fuhrman, S. H. (2003). *Redesigning accountability systems in education.* Retrieved on April 23, 2011 from Consortium for Policy Research in Education website: http://www.cpre.org/images/stories/cpre_pdfs/rb38.pdf

Goertz, M. E., & Massell, D. (2005). *Holding high hopes: How high schools respond to state accountability policies.* Retrieved on April 23, 2011 from Consortium for Policy Research in Education website: http://www.cpre.org/images/stories/cpre_pdfs/rb27.pdf

Goodwin, B. (2011). *Simply better.* Alexandria, VA: Association for Supervision and Staff Development.

Hanushek, E. A., & Raymond, M. E., (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis, 24*, 297-327.

Harris, D. N. (2010). Clear away the smoke and mirrors of value-added. *Phi Delta Kappan*, *91*, 66-69.

Harris, D. N., & Herrington, C. D. (2006). Accountability, standards, and the growing achievement gap: Lessons from the past half-century. *American Journal of Education, 112,* 209-238.

Hattie, J. A. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York, NY: Routledge.

Heim, M. (1996) *Accountability in education: A primer for school leaders.* A monograph prepared for Hawaii School Leadership Academy. Retrieved from ERIC Database. (ED444267)

Hershberg, T., Simon, V., & Lea-Kruger, B. (2004). *Measuring what matters: How value added assessment can be used to drive learning gains*. Retrieved on July 23, 2011 from the Center for Greater Philadelphia at the website: http://www.cgp.upenn.edu/pdf/measuring_what_matters.pdf

Hewitt, T. W. (2008). Speculations on *A Nation at Risk*: Illusions and realities. *Phi Delta Kappan, 89*, 575-579.

Hochberg, E. D., & Desimone, L. M. (2010). Professional development in the accountability context: Building capacity to achieve standards. *Educational Psychologist, 45*, 89-106.

Hunt, T. C. (2005). Education reforms: Lessons from history. *Phi Delta Kappan, 87*, 84-89.

Johnson, B. (2001). Toward a new classification of non-experimental quantitative research. *Educational Researcher, 30*(2), 3-13.

Kirst, M. (1990). *Accountability: Implications for state and local policy makers*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, Information Services. Retrieved from ERIC Database. (ED318804)

Lee, J. L. (2006). Input-guarantee versus performance-guarantee approaches to school accountability: Cross-state comparisons of policies, resources, and outcomes. *Peabody Journal of Education, 81*(4), 43-64.

Lee, J. L., & Wong, K. K. (2004). The impact of accountability on racial and socioeconomic equity: Considering both school resources and achievement outcomes. *American Education Research Journal, 41*, 797-832.

Linn, R. L. (2008). Methodological issues in achieving school accountability. *Journal of Curriculum Studies, 40*, 699-711.

Lissitz, B., & Doran, H. (2009). *Modeling growth and accountability.* Retrieved from: dpi.wi.gov/oea/pdf/introgrowth.pdf

Lockwood, J . R., McCaffrey, D. F., Hamilton, L. S., Stecher, B., Le, V., & Martinez, J. F. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement, 44*(1), 47-67.

Manno, B. V. (2004). Chartering and the idea of accountability consequences: Adding performance value to schooling. *Journal Of Education*, *185*(3), 27-40. Retrieved from ERIC database. (EJ774575)

Marzano, R. J. (2003a). *What works in schools: Facilitator's guide.* Alexandria, VA: Association for Supervision and Staff Development.

Marzano, R. J. (2003b). *What works in schools: Translating research into action.* Alexandria, VA: Association for Supervision and Staff Development.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND Corporation.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal for Educational and Behavioral Statistics, 29*(1), 67-101.

McClure, P. (2005). *School improvement under No Child Left Behind.* Retrieved on July, 24, 2011from the Center for American Progress website: http://www.american progress.org/kf/mcclure3-03-2005.pdf

Meyer, R. H. (2000). *Value-added indicators: A powerful tool for evaluating science and mathematics programs and policies.* Retrieved on July 23, 2011 from the Wisconsin Center for Educational Research website: http://www.wceruw.com /archive/nise/Publications/Briefs/Vol_3_No_3/Vol%203,%20No%203.pdf

Murnane, R. J. (2007). *Improving the education of children living in poverty*. Retrieved on September 11, 2011 from the Future of Children website: http://futureofchildren.org/futureofchildren/publications/docs/17_02_08.pdf

Newmann, F. M., King, M. B., & Rigdon, M. (1997). Accountability and school performance: Implications from restructuring schools. *Harvard Educational Review, 67*(1), 41.

No Child Left Behind Act of 2001. (2002). Retrieved from U.S. Department of Education website: http://www2.ed.gov/policy/elsec/leg/esea02/index.html

Olsen, B., & Sexton, D. (2009). Threat rigidity, school reform, and how teachers view their work inside current education policy contexts. *American Educational Research Journal, 46*(1), 9-44.

O'Shea, M. R. (2005). *From standards to success.* Alexandria, VA: Association for Supervision and Staff Development.

PVAAS Information: History, Intent and Timeline for Implementation (2011). Retrieved from the Pennsylvania Department of Education Website: http://www.portal. state.pa.us/portal/http;/ /www.portal.state.pa.us;80/portal/server.pt/gateway /PTARGS_0_123031_1039888_0_0_18/PVAAS_Info_1-31-11.pdf

PVAAS Statewide Core team for PDE. (2011). *Pennsylvania value-added assessment system (PVAAS): Guide for districts and schools on PVAAS reporting.* Retrieved from https://pvaas.sas. com/evaas/signin.jsf

PVAAS: Statewide Plan Pennsylvania Statewide Plan for PVAAS (2006). Retrieved from the Pennsylvania Department of Education website: http://www.portal .state.pa.us/portal/ server.pt?open= 18&objID=354364&mode=2

Raudenbush, S.W. (2004). *Schooling, statistics, and poverty: Can we measure school improvement?*. Retrieved on July 24, 2011 from the Educational Testing Service website: http://www.ets.org/Media/Education_Topics/pdf/angoff9.pdf

Raudenbush, S.W., & Willms, D. J. (1995). The estimation of teacher effects. *The Journal of Educational and Behavioral Statistics, 20*, 307-335.

Reeves, D. B. (2005). *Accountability on action: A blueprint for learning organizations.* Englewood, CO: Lead and Learn Press.

Rosenshine, B. (2003). High stakes testing: Another analysis. *Education Policy Analysis Archives, 11*(24), 1-8. Retrieved 9/11/2011 from http://epaa.asu. edu/epaa/ v11n24/

Sanders, W. L. (2000). Value-added assessments from student achievement data: Opportunities and hurdles. *Journal of Personnel Evaluation in Education, 14*, 329-339.

Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee value-added assessment system (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education, 12*, 247-256.

Schmoker, M. (1999). *Results: The key to continuous school improvement* (2nd ed.). Alexandria, VA: Association for Supervision and Staff Development.

Schmoker, M. (2006). *Results now: How we can achieve unprecedented improvements in teaching and learning.* Alexandria, VA: Association for Supervision and Staff Development.

Sirotnik, K. A. (2004). *Holding accountability accountable: What ought to matter in public education*. New York, NY: Teachers College Press.

Song, J., & Felch, J. (2011, May 7). Times updates and expands value-added ratings for Los Angeles elementary school teachers. *Los Angeles Times.* Retrieved from http://www.latimes.com/news/local/la-me-value-added-20110508,0,930050.story

Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of Educational Change, 7*(1), 221-258.

Superfine, B. M. (2005). The politics of accountability: The rise and fall of goals 2000. *American Journal of Education, 112*(1), 10-43.

Topping, K. J., & Sanders, W. L. (2000). Teacher effectiveness and computer assessment of reading: Relating value-added and learning information systems. *School Effectiveness and School Improvement, 11*(3), 305-337.

Tosh, K., & Edwards, C. J. (2009). The new Title I: The changing landscape of accountability. Washington, DC: Thompson Publishing.

Troen, V., & Boles, K. C. (2012). *The power of teacher teams.* Thousand Oaks, CA: Sage.

U.S. Department of Education, Institute of Education Sciences. (1999). *Highlights from trends in mathematics and science study* (NCES Publication No: 1999081). Retrieved from http://nces.ed.gov/pubs99/1999081.pdf

Vanhuysse, P., & Sulitzeanu-Kenan, R. (2009). Teacher's PAT? Multiple-role principal–agent theory, education politics, and bureaucrat power. *Critical Studies in Education,* 5(2), 129-144.

Wescott, J. P. (1972). *Accountability: For whom, to whom, for what?* [Speech presented at the 104th Annual Convention of American Association of School Administrators]. Atlantic City NJ. Retrieved from ERIC Database. (ED059964)

White, S., & Smith, R. L. (2010). *School improvement for the next generation.* Bloomington, IN: Solution Tree Press.

Wiliam, D. (2010). Standardized testing and school accountability. *Educational Psychologist,45*(2), 107-122.

Willms, D. J., & Raudenbush, S. W. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement, 26*(3), 209-232.

Wohlstetter, P. (1991). Accountability mechanisms for state and education reform: Some organizational alternatives. *Educational Evaluation and Policy Analysis, 13*(1), 31-48.

Wohlstetter, P., Datnow, A., & Park, V. (2008). Creating a system for data-driven decision-making: Applying the principal-agent framework. *School Effectiveness and School Improvement*, *19*(3), 239-259.

Wong, K. K. (2003). Emerging issues in educational reform in Japan. *Peabody Journal of Education, 78*(4), 242-248.

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

APPENDIX A

AGREEMENT TO CONDUCT A STUDY USING UNIT LEVEL PSSA DATA FROM THE PENNSYLVANIA Department of Education; Bureau of Assessment and Accountability

This agreement is made and entered into by and between the Pennsylvania Department of Education (PDE), Bureau of Assessment and Accountability (BAA) and _____ (the "Researcher").

Whereas, PDE administers a statewide repository of PSSA Data including student demographic information, and

Whereas, these data may be made available to researchers in a manner consistent with the Family Educational Rights and Privacy Act of 1974 (FERPA); and

Whereas, the Researcher is conducting a study and has asked PDE for access to de-identified PSSA data. The research study is described in the Attachments, which are incorporated herein and attached hereto.

Now therefore, for good and valuable consideration, the receipt and sufficiency of which are hereby acknowledged, PDE and the Researcher agree as follows:

1. The Researcher will be given access to confidential student information for the limited purpose of conducting the study described in the Attachments. The Researcher understands and agrees that any unauthorized disclosure of confidential student information is illegal as provided in the Family Education Rights and Privacy Act of 1974 (FERPA) and in the implementing federal regulations found in 34 CFR, Part 99. The Researcher agrees to protect any personal characteristics of a student that could make the student's identity traceable by using a minimum confidentiality *n* of10. Any data sets or output reports that the Researcher, or its authorized agents, may generate using confidential data are to be protected. The Researcher agrees that any data analysis or report will not be disclosed to any party without the consent of PDE.

The Researcher agrees that only the following persons will have access to the data:

_____
_____
_____

The Researcher agrees that a confidentiality agreement in the form attached hereto shall be signed by each person named above. No other person may have access to the data unless PDE agrees, by written amendment, to permit such access. A confidentiality agreement must be signed by each additional person who is given access to the data. The Researcher understands that the failure to observe these restrictions may expose the Researcher to liability and may result in the termination of this Agreement.

More>>

2. The Researcher agrees that access to the data shall be limited to the amount of time necessary to complete the study and that, once the study has been completed, the Researcher shall destroy or return the data in a manner agreeable to PDE.

3. The Researcher shall maintain all PSSA education records received from PDE separate from all other data files on a secure password protected computer and shall not provide copies or extracts of the PSSA data to employees or subcontractors not associated with the research project.

4. The Researcher shall provide PDE with periodic status reports outlining the progress of the research study at a frequency to be determined by PDE.

5. The Researcher shall not disclose PSSA data summaries or otherwise release data or reports relating to any individual student.

6. The Researcher shall not disclose PSSA data summaries or otherwise release data or reports relating to any group or category of student in the data file without ensuring the confidentiality of students in that group. Any publication or report produced using education records from BAA should include only aggregate summaries and no personally identifiable information or other information that could lead to the identification of any student.

7. The Researcher shall destroy or return all student education records from BAA, in a manner acceptable to PDE, when no longer needed for the particular research project as required under FERPA regulations Section 99.35(b)(2).

8. The Researcher shall not use the student education records to contact individuals who are data subjects.

9. The Researcher shall provide PDE with one electronic and one paper copy of the final versions of all reports and related documents prepared using education records from BAA and shall notify PDE in writing of any differences between reports of PDE data, calculation models, performance level percentages or any other PDE methodology and those used in the report at least 30 days prior to any publication using the data. Upon request of PDE, The Researcher shall include in its final report, and in any copy, or derivation of such report, disclaimer language, in form and substance acceptable to PDE, which shall disclaim PDE's responsibility for the report and, if requested by PDE, identify any differences between reports of PDE data, calculation models, performance level percentages or any other PDE methodology and those used in the report. PDE reserves the right to distribute and otherwise utilize any final reports and related documents as it wishes, in sum or in part.

10. The Researcher agrees to the following additional requirements:

_____

_____

_____

11. PDE reserves the right to charge a reasonable fee for the use of data by researchers to help offset the state's costs of collecting and storing the data when such researchers are not affiliated with Institutions of Postsecondary Education or Local Education Agencies that have supplied education records for the purpose of the PSSA.

More>>

12. PDE reserves the right to use, make copies of and distribute the final version of reports or research studies created by Researcher using education record data from BAA.

13. Either PDE or the Researcher may terminate this Agreement with ten (10) days written notice and all data in the possession of Researcher at that time will be destroyed or returned within ten (10) days of the date of termination.

14. This Agreement and attachments hereto constitute the entire agreement between the parties. No agent, representative, employee, or officer of either the Commonwealth or the Researcher has authority to make, or has made, any statement, agreement or representation, oral or written, in connection with this Agreement, which in any way can be deemed to modify, add to or detract from, or otherwise change or alter its terms and conditions. No negotiations between the parties, nor any custom or usage, shall be permitted to modify or contradict any of the terms and conditions of the Agreement. No modifications, alterations, changes, or waiver to this Agreement or any of its terms shall be valid or binding unless accomplished by a written amendment signed by both parties.

15. Absent notice to the contrary in writing, all communications to PDE shall be sent to:

Bureau of Assessment and Accountability
Division of Performance Analysis and Reporting
Pennsylvania Department of Education
333 Market Street
Harrisburg, PA 17126

Absent notice to the contrary in writing, all communications to the Researcher shall be sent to:

_____
_____
_____

16. The rights, duties, obligations and interests of the parties set out herein shall not be assigned or transferred.

17. Nothing herein shall be construed to create any personal liability on the part of any officer or agent of either party hereto.

18. Nothing in this agreement shall be construed to violate any provision of the laws and/or regulations of the United States of America or the Commonwealth of Pennsylvania, and all acts done hereunder shall be done in such manner as may conform thereto. If any word, phrase, clause, paragraph, sentence, part, portion, or provision of this agreement or the application thereof to any person or circumstance is held to be invalid, the remainder of this agreement shall nevertheless be valid, and the parties hereby declare that this agreement would have been executed without such invalid word, phrase clause, paragraph, sentence, part, portion, or provision. All of the terms and provisions of this agreement are to be construed to effectuate the purpose, powers, rights, functions, and authorities herein set forth.

19. The parties rights and obligations under paragraphs 5, 6, 8, 12 and 18 shall survive the termination of this agreement.

More>>

20. This Agreement shall be governed by, interpreted, and enforced in accordance with the laws of the Commonwealth of Pennsylvania (without regard to conflict of laws provisions) and the decisions of the Pennsylvania courts. Researcher consents to the jurisdiction of any court of the Commonwealth of Pennsylvania and any federal courts in Pennsylvania, waiving any claim or defense that such forum is not convenient or proper. Researcher agrees that any such court shall have personal or subject matter jurisdiction over it, and consents to service of process in any manner authorized by Pennsylvania law.

In witness whereof, the Pennsylvania Department of Education and Researcher have executed this agreement through their authorized representatives to be effective this _____ day of_____, 20__ and to end upon the earlier of the conclusion of the study or termination under paragraph thirteen (13).

Pennsylvania Department of Education

By_____
[NAME]
[TITLE]

(NAME OF THE ORGANIZATION OR INDIVIDUAL)

By_____

Printed Name:_____

Title: _____

More>>

ACCESS AGREEMENT

Under the terms of the Agreement between the Pennsylvania Department of Education (the "PDE") and _____(" "), I am being provided with access to confidential student information.

I understand that any unauthorized disclosure of confidential student information is illegal as provided in the Family Educational Rights and Privacy Act of 1974 (FERPA) and in the implementing federal regulations found at 34 CFR, Part 99. I understand that any personal characteristics of a student that could make the student's identity traceable shall be protected.

In addition, I understand that any data sets or output reports that I may generate using confidential data shall be protected. I shall not distribute to any unauthorized person any data sets or reports that I have access to or may generate using confidential data. I shall not disclose any data analysis or report, except as provided for in the Agreement between PDE and my employer or contractor, and only after an appropriate confidentiality agreement is executed by same. I understand that I am responsible for any telephone or computer transactions or costs of same performed as a result of access authorized by use of signon/password(s).

_____ _____

Signature Date

_____

Printed Name

More>>

ATTACHMENTS

Attachment A: Researcher Information
The researcher shall provide a curriculum vitae for each of the researchers involved in the project and identify the Principal Investigator.

Attachment B: Project Information
The researcher shall provide a brief description of the research project including summary of the topic to be researched and list of research questions. The researcher should provide information for any/all non-BAA data that will be used as part of this research.

Attachment C: Data Element Crosswalk
The researcher shall provide a specific listing of the items of data being requested and provide a rationale for each data element.

Attachment D: Non-data Information Requests
The researcher shall provide a list of any non-data information requests such as but not limited to letters of support or subject matter expertise (provide estimated number of hours).

Attachment E: Research Methodology
The researcher shall provide a description of the research methods that will be employed in this study and demonstrate how the data and methods are suitable to answer the research questions.

Attachment F: Timeline Requirements
The researcher shall provide a detailed timeline of the entire research project. Timeline must include Data collection, analysis, report writing, review and publication. Please provide anticipated date by which any publications will be made public. PDE's response to data access requests will depend upon the timing and nature of the request. Likewise, the amount of time it takes PDE to transmit the requested data will also depend upon the timing and nature of the request.

Attachment G: Project Budget
The researcher shall provide a proposed budget for the research project and identify all sources and amounts of project funding that has been secured or is pending.

Attachment H: Statement of Benefits
The researcher shall demonstrate the benefit (potential for improving instruction and educational outcomes) of the proposed research and how PA can use the research in its final form. Note that demonstration of research benefits does not imply approval of the proposal.

Attachment I: Evidence of Compliance
The Researcher shall provide a description of the procedures that will be taken to ensure

More>>

that the requested data is handled, stored, maintained, analyzed and reported in a manner that is in compliance with Pennsylvania Department of Education's "Student Data Access and Use Policy" and the conditions agreed to in the access agreement above.

Attachment J: Evidence that The Researcher is Working as an "Authorized Representative" of The Pennsylvania Department Of Education.
The Researcher shall outline the facts that demonstrate that the researcher is working as a contractor, or other party under the direct control of PDE, to audit or evaluate Federal or State supported education programs, or for the enforcement of or compliance with Federal legal requirements that relate to those programs.

APPENDIX B

Frequency Distribution for 2010 Mathematics PVAAS

		Frequency	Percent	Cumulative Frequency	Cumulative Percent
Valid	-10.6	1	0.2	1	0.2
	-10.0	1	0.2	2	0.5
	-9.5	1	0.2	3	0.7
	-9.2	2	0.5	5	1.2
	-9.1	1	0.2	6	1.4
	-9.0	1	0.2	7	1.6
	-8.8	1	0.2	8	1.9
	-7.6	1	0.2	9	2.1
	-7.3	2	0.5	11	2.6
	-7.1	3	0.7	14	3.3
	-7.0	4	0.9	18	4.2
	-6.9	1	0.2	19	4.5
	-6.8	2	0.5	21	4.9
	-6.7	2	0.5	23	5.4
	-6.6	3	0.7	26	6.1
	-6.5	1	0.2	27	6.3
	-6.4	1	0.2	28	6.6
	-6.3	1	0.2	29	6.8
	-6.1	2	0.5	31	7.3
	-6.0	1	0.2	32	7.5
	-5.9	2	0.5	34	8.0
	-5.8	4	0.9	38	8.9
	-5.7	1	0.2	39	9.2
	-5.6	3	0.7	42	9.9
	-5.5	1	0.2	43	10.1
	-5.4	1	0.2	44	10.3
	-5.3	3	0.7	47	11.0
	-5.2	4	0.9	51	12.0
	-5.1	2	0.5	53	12.4
	-5.0	1	0.2	54	12.7
	-4.9	2	0.5	56	13.1
	-4.8	1	0.2	57	13.4

More>>

2010 PVAAS Mathematics (continued)

		Frequency	Percent	Cumulative Frequency	Cumulative Percent
Valid	-4.5	1	0.2	58	13.6
	-4.4	1	0.2	59	13.8
	-4.3	2	0.5	61	14.3
	-4.1	2	0.5	63	14.8
	-4.0	5	1.2	68	16.0
	-3.9	6	1.4	74	17.4
	-3.8	2	0.5	76	17.8
	-3.7	5	1.2	81	19.0
	-3.6	2	0.5	83	19.5
	-3.5	2	0.5	85	20.0
	-3.4	3	0.7	88	20.7
	-3.3	3	0.7	91	21.4
	-3.2	3	0.7	94	22.1
	-3.1	3	0.7	97	22.8
	-3.0	6	1.4	103	24.2
	-2.9	2	0.5	105	24.6
	-2.8	2	0.5	107	25.1
	-2.7	4	0.9	111	26.1
	-2.6	3	0.7	114	26.8
	-2.5	6	1.4	120	28.2
	-2.4	3	0.7	123	28.9
	-2.3	3	0.7	126	29.6
	-2.2	5	1.2	131	30.8
	-2.1	3	0.7	134	31.5
	-2.0	3	0.7	137	32.2
	-1.9	6	1.4	143	33.6
	-1.8	1	0.2	144	33.8
	-1.7	3	0.7	147	34.5
	-1.6	4	0.9	151	35.4
	-1.5	10	2.3	161	37.8
	-1.4	1	0.2	162	38.0
	-1.3	1	0.2	163	38.3

More>>

2010 PVAAS Mathematics (continued)

		Frequency	Percent	Cumulative Frequency	Cumulative Percent
Valid	-1.2	4	0.9	167	39.2
	-1.1	4	0.9	171	40.1
	-1.0	2	0.5	173	40.6
	-0.9	6	1.4	179	42.0
	-0.8	3	0.7	182	42.7
	-0.7	3	0.7	185	43.4
	-0.6	3	0.7	188	44.1
	-0.5	5	1.2	193	45.3
	-0.4	3	0.7	196	46.0
	-0.3	6	1.4	202	47.4
	-0.2	8	1.9	210	49.3
	-0.1	7	1.6	217	50.9
	0.0	3	0.7	220	51.6
	0.1	5	1.2	225	52.8
	0.2	1	0.2	226	53.1
	0.3	2	0.5	228	53.5
	0.4	2	0.5	230	54.0
	0.5	1	0.2	231	54.2
	0.6	3	0.7	234	54.9
	0.7	7	1.6	241	56.6
	0.8	3	0.7	244	57.3
	0.9	5	1.2	249	58.5
	1.0	4	0.9	253	59.4
	1.1	4	0.9	257	60.3
	1.2	5	1.2	262	61.5
	1.3	2	0.5	264	62.0
	1.4	3	0.7	267	62.7
	1.5	2	0.5	269	63.1
	1.6	7	1.6	276	64.8
	1.7	2	0.5	278	65.3
	1.8	6	1.4	284	66.7
	1.9	3	0.7	287	67.4

More>>

2010 PVAAS Mathematics (continued)

		Frequency	Percent	Cumulative Frequency	Cumulative Percent
Valid	2.0	4	0.9	291	68.3
	2.1	3	0.7	294	69.0
	2.2	2	0.5	296	69.5
	2.3	6	1.4	302	70.9
	2.4	4	0.9	306	71.8
	2.5	4	0.9	310	72.8
	2.6	2	0.5	312	73.2
	2.7	4	0.9	316	74.2
	2.8	1	0.2	317	74.4
	2.9	1	0.2	318	74.6
	3.0	3	0.7	321	75.4
	3.1	1	0.2	322	75.6
	3.2	3	0.7	325	76.3
	3.3	2	0.5	327	76.8
	3.4	2	0.5	329	77.2
	3.5	3	0.7	332	77.9
	3.6	3	0.7	335	78.6
	3.7	1	0.2	336	78.9
	4.0	3	0.7	339	79.6
	4.1	4	0.9	343	80.5
	4.2	6	1.4	349	81.9
	4.3	3	0.7	352	82.6
	4.4	3	0.7	355	83.3
	4.5	1	0.2	356	83.6
	4.7	3	0.7	359	84.3
	4.8	2	0.5	361	84.7
	4.9	3	0.7	364	85.4
	5.0	2	0.5	366	85.9
	5.1	2	0.5	368	86.4
	5.2	1	0.2	369	86.6
	5.4	1	0.2	370	86.9
	5.5	2	0.5	372	87.3

More>>

2010 PVAAS Mathematics (continued)

		Frequency	Percent	Cumulative Frequency	Cumulative Percent
Valid	5.6	1	.2	373	87.6
	5.7	2	.5	375	88.0
	5.8	2	.5	377	88.5
	6.0	2	.5	379	89.0
	6.1	1	.2	380	89.2
	6.2	2	.5	382	89.7
	6.3	3	.7	385	90.4
	6.4	1	.2	386	90.6
	6.5	2	.5	388	91.1
	6.6	3	.7	391	91.8
	6.7	1	.2	392	92.0
	7.0	1	.2	393	92.3
	7.1	1	.2	394	92.5
	7.2	1	.2	395	92.7
	7.3	1	.2	396	93.0
	7.5	2	.5	398	93.4
	7.7	2	.5	400	93.9
	7.9	1	.2	401	94.1
	8.0	2	.5	403	94.6
	8.1	2	.5	405	95.1
	8.2	1	.2	406	95.3
	8.5	1	.2	407	95.5
	8.7	1	.2	408	95.8
	8.8	2	.5	410	96.2
	8.9	1	.2	411	96.5
	9.0	1	.2	412	96.7
	9.1	1	.2	413	96.9
	9.5	1	.2	414	97.2
	9.7	1	.2	415	97.4
	10.1	1	.2	416	97.7
	10.3	1	.2	417	97.9
	10.4	1	.2	418	98.1

More>>

2010 PVAAS Mathematics (continued)

		Frequency	Percent	Cumulative Frequency	Cumulative Percent
Valid	11.2	1	.2	419	98.4
	11.6	1	.2	420	98.6
	11.9	1	.2	421	98.8
	12.0	1	.2	422	99.1
	13.1	1	.2	423	99.3
	15.4	1	.2	424	99.5
	16.4	1	.2	425	99.8
	20.0	1	.2	426	100.0
	Total	426	100.0		

APPENDIX C

Frequency Distribution for 2010 Reading PVAAS

		Frequency	Percent	Cumulative Frequency	Cumulative Percent
Valid	-10.5	1	.2	1	.2
	-9.4	1	.2	2	.5
	-9.2	1	.2	3	.7
	-9.1	1	.2	4	.9
	-8.6	1	.2	5	1.2
	-8.1	1	.2	6	1.4
	-8.0	1	.2	7	1.6
	-7.4	1	.2	8	1.9
	-7.2	1	.2	9	2.1
	-7.0	1	.2	10	2.3
	-6.9	1	.2	11	2.6
	-6.8	1	.2	12	2.8
	-6.6	2	.5	14	3.3
	-6.3	1	.2	15	3.5
	-5.9	1	.2	16	3.8
	-5.8	1	.2	17	4.0
	-5.7	2	.5	19	4.5
	-5.5	1	.2	20	4.7
	-5.4	3	.7	23	5.4
	-5.1	2	.5	25	5.9
	-5.0	1	.2	26	6.1
	-4.8	1	.2	27	6.3
	-4.7	2	.5	29	6.8
	-4.6	3	.7	32	7.5
	-4.5	1	.2	33	7.7
	-4.4	5	1.2	38	8.9
	-4.3	3	.7	41	9.6
	-4.2	5	1.2	46	10.8
	-4.1	1	.2	47	11.0
	-4.0	5	1.2	52	12.2
	-3.9	1	.2	53	12.4
	-3.8	3	.7	56	13.1

More>>

2010 PVAAS Reading (Continued)

		Frequency	Percent	Cumulative Frequency	Cumulative Percent
Valid	-3.7	2	.5	58	13.6
	-3.6	1	.2	59	13.8
	-3.5	3	.7	62	14.6
	-3.4	1	.2	63	14.8
	-3.3	5	1.2	68	16.0
	-3.2	5	1.2	73	17.1
	-3.1	2	.5	75	17.6
	-3.0	4	.9	79	18.5
	-2.8	7	1.6	86	20.2
	-2.7	1	.2	87	20.4
	-2.6	1	.2	88	20.7
	-2.5	3	.7	91	21.4
	-2.4	3	.7	94	22.1
	-2.3	1	.2	95	22.3
	-2.2	4	.9	99	23.2
	-2.1	5	1.2	104	24.4
	-2.0	4	.9	108	25.4
	-1.9	10	2.3	118	27.7
	-1.8	3	.7	121	28.4
	-1.7	8	1.9	129	30.3
	-1.6	3	.7	132	31.0
	-1.5	8	1.9	140	32.9
	-1.4	7	1.6	147	34.5
	-1.3	3	.7	150	35.2
	-1.2	5	1.2	155	36.4
	-1.1	1	.2	156	36.6
	-1.0	8	1.9	164	38.5
	-.9	6	1.4	170	39.9
	-.8	6	1.4	176	41.3
	-.7	6	1.4	182	42.7
	-.6	8	1.9	190	44.6
	-.5	7	1.6	197	46.2

More>>

2010 PVAAS Reading (Continued)

		Frequency	Percent	Cumulative Frequency	Cumulative Percent
Valid	-.4	2	.5	199	46.7
	-.3	3	.7	202	47.4
	-.2	6	1.4	208	48.8
	-.1	4	.9	212	49.8
	.0	5	1.2	217	50.9
	.1	6	1.4	223	52.3
	.2	5	1.2	228	53.5
	.3	6	1.4	234	54.9
	.4	7	1.6	241	56.6
	.5	6	1.4	247	58.0
	.6	7	1.6	254	59.6
	.7	6	1.4	260	61.0
	.8	4	.9	264	62.0
	.9	4	.9	268	62.9
	1.0	1	.2	269	63.1
	1.1	5	1.2	274	64.3
	1.2	4	.9	278	65.3
	1.3	9	2.1	287	67.4
	1.4	4	.9	291	68.3
	1.5	5	1.2	296	69.5
	1.6	3	.7	299	70.2
	1.7	4	.9	303	71.1
	1.8	3	.7	306	71.8
	1.9	6	1.4	312	73.2
	2.0	2	.5	314	73.7
	2.1	3	.7	317	74.4
	2.2	4	.9	321	75.4
	2.3	5	1.2	326	76.5
	2.4	4	.9	330	77.5
	2.5	6	1.4	336	78.9
	2.6	3	.7	339	79.6
	2.7	4	.9	343	80.5

More>>

2010 PVAAS Reading (Continued)

		Frequency	Percent	Cumulative Frequency	Cumulative Percent
Valid	2.8	2	.5	345	81.0
	3.0	5	1.2	350	82.2
	3.1	2	.5	352	82.6
	3.2	2	.5	354	83.1
	3.3	3	.7	357	83.8
	3.4	2	.5	359	84.3
	3.5	2	.5	361	84.7
	3.6	3	.7	364	85.4
	3.9	1	.2	365	85.7
	4.0	3	.7	368	86.4
	4.1	2	.5	370	86.9
	4.3	3	.7	373	87.6
	4.4	3	.7	376	88.3
	4.5	1	.2	377	88.5
	4.6	3	.7	380	89.2
	4.7	3	.7	383	89.9
	4.8	1	.2	384	90.1
	4.9	1	.2	385	90.4
	5.1	1	.2	386	90.6
	5.2	1	.2	387	90.8
	5.3	3	.7	390	91.5
	5.5	2	.5	392	92.0
	5.6	2	.5	394	92.5
	5.7	1	.2	395	92.7
	5.8	2	.5	397	93.2
	5.9	3	.7	400	93.9
	6.0	3	.7	403	94.6
	6.1	1	.2	404	94.8
	6.3	2	.5	406	95.3
	6.5	2	.5	408	95.8
	6.7	1	.2	409	96.0
	6.9	1	.2	410	96.2

More>>

2010 PVAAS Reading (Continued)

		Frequency	Percent	Cumulative Frequency	Cumulative Percent
Valid	7.0	1	.2	411	96.5
	7.1	2	.5	413	96.9
	7.5	2	.5	415	97.4
	8.6	1	.2	416	97.7
	9.4	2	.5	418	98.1
	12.4	1	.2	419	98.4
	15.0	7	1.6	426	100.0
	Total	426	100.0		

APPENDIX D

Participating Pennsylvania High Schools (continued)

High School Number	2010 PVAAS Math	2010 PVAAS Reading	2010 SIC A	2010 SIC B
High School 1	-9.2	-4.0	A1	B1
High School 2	-9.2	-2.5	A1	B1
High School 3	-7.0	-4.8	A1	B1
High School 4	-6.9	-6.8	A1	B1
High School 5	-6.8	0.7	A1	B1
High School 6	-6.3	-4.2	A1	B1
High School 7	-6.0	-0.5	A1	B1
High School 8	-5.8	-3.3	A1	B1
High School 9	-5.2	-1.9	A1	B1
High School 10	-5.2	-3.7	A1	B1
High School 11	-5.2	2.0	A1	B1
High School 12	-4.8	-5.9	A1	B1
High School 13	-4.4	-7.4	A1	B1
High School 14	-4.1	-4.0	A1	B1
High School 15	-4.0	5.6	A1	B1
High School 16	-4.0	-1.5	A1	B1
High School 17	-4.0	-2.3	A1	B1
High School 18	-3.9	-1.0	A1	B1
High School 19	-3.7	-2.0	A1	B1
High School 20	-3.7	1.9	A1	B1
High School 21	-3.7	2.5	A1	B1
High School 22	-3.6	-0.1	A1	B1
High School 23	-3.6	-1.9	A1	B1
High School 24	-3.5	-1.3	A1	B1
High School 25	-3.4	-0.6	A1	B1
High School 26	-3.4	-4.6	A1	B1
High School 27	-3.4	-1.5	A1	B1
High School 28	-3.3	-1.0	A1	B1
High School 29	-3.1	-4.1	A1	B1
High School 30	-3.1	-0.8	A1	B1
High School 31	-3.0	1.3	A1	B1
High School 32	-3.0	-2.8	A1	B1
High School 33	-3.0	-2.8	A1	B1
High School 34	-2.8	1.1	A1	B1
High School 35	-2.6	-1.9	A1	B1
High School 36	-2.5	-2.0	A1	B1
High School 37	-2.3	-1.4	A1	B1
High School 38	-2.2	-1.6	A1	B1
High School 39	-2.2	-1.7	A1	B1
High School 40	-2.1	-2.0	A1	B1
High School 41	-1.9	-3.2	A1	B1
High School 42	-1.8	-0.7	A1	B1

Participating Pennsylvania High Schools (continued)

High School Number	2010 PVAAS Math	2010 PVAAS Reading	2010 SIC A	2010 SIC B
High School 43	-1.5	-1.1	A1	B1
High School 44	-1.2	2.6	A1	B1
High School 45	-1.2	-0.7	A1	B1
High School 46	-1.1	0.4	A1	B1
High School 47	-1.1	-0.2	A1	B1
High School 48	-1.0	-0.5	A1	B1
High School 49	-0.9	-2.6	A1	B1
High School 50	-0.9	-1.5	A1	B1
High School 51	-0.9	0.2	A1	B1
High School 52	-0.7	1.3	A1	B1
High School 53	-0.6	5.5	A1	B1
High School 54	-0.6	-2.4	A1	B1
High School 55	-0.3	0.1	A1	B1
High School 56	-0.2	2.8	A1	B1
High School 57	-0.2	0.2	A1	B1
High School 58	-0.2	1.5	A1	B1
High School 59	-0.2	-3.2	A1	B1
High School 60	-0.1	3.0	A1	B1
High School 61	-0.1	4.0	A1	B1
High School 62	-0.1	2.7	A1	B1
High School 63	0.0	-1.0	A1	B1
High School 64	0.0	-0.5	A1	B1
High School 65	0.1	-0.9	A1	B1
High School 66	0.1	0.4	A1	B1
High School 67	0.1	-0.6	A1	B1
High School 68	0.3	4.9	A1	B1
High School 69	0.5	3.0	A1	B1
High School 70	0.6	-1.9	A1	B1
High School 71	0.7	-2.1	A1	B1
High School 72	0.8	-1.9	A1	B1
High School 73	1.0	3.3	A1	B1
High School 74	1.0	-2.5	A1	B1
High School 75	1.1	2.1	A1	B1
High School 76	1.2	5.8	A1	B1
High School 77	1.7	2.4	A1	B1
High School 78	2.0	2.1	A1	B1
High School 79	2.0	0.3	A1	B1
High School 80	2.2	-0.5	A1	B1
High School 81	2.2	-0.6	A1	B1
High School 82	2.3	1.5	A1	B1
High School 83	2.4	6.3	A1	B1
High School 84	2.4	0.4	A1	B1
High School 85	2.5	0.5	A1	B1
High School 86	2.5	2.7	A1	B1
High School 87	2.6	-0.7	A1	B1
High School 88	2.7	-0.2	A1	B1
High School 89	3.0	3.5	A1	B1

Participating Pennsylvania High Schools (continued)

High School Number	2010 PVAAS Math	2010 PVAAS Reading	2010 SIC A	2010 SIC B
High School 90	3.5	3.3	A1	B1
High School 91	3.5	0.1	A1	B1
High School 92	3.6	1.4	A1	B1
High School 93	3.6	1.2	A1	B1
High School 94	3.7	4.3	A1	B1
High School 95	4.0	1.5	A1	B1
High School 96	4.0	2.7	A1	B1
High School 97	4.3	0.2	A1	B1
High School 98	4.4	0.9	A1	B1
High School 99	4.7	-1.2	A1	B1
High School 100	4.8	-1.9	A1	B1
High School 101	5.0	5.8	A1	B1
High School 102	5.1	1.8	A1	B1
High School 103	5.7	4.4	A1	B1
High School 104	6.3	2.3	A1	B1
High School 105	6.4	-0.2	A1	B1
High School 106	6.5	0.8	A1	B1
High School 107	6.6	0.3	A1	B1
High School 108	7.2	7.1	A1	B1
High School 109	8.1	5.9	A1	B1
High School 110	10.3	2.5	A1	B1
High School 111	15.4	3.9	A1	B1
High School 112	-7.3	-4.2	A10	B7
High School 113	-7.0	-4.3	A10	B7
High School 114	-4.9	-4.4	A10	B7
High School 115	-4.5	0.3	A10	B7
High School 116	-1.6	-1.3	A10	B7
High School 117	-1.5	-0.9	A10	B7
High School 118	-0.7	0.4	A10	B7
High School 119	2.5	2.7	A10	B7
High School 120	10.4	5.6	A10	B7
High School 121	-9.1	-4.2	A11	B8
High School 122	-9.0	-5.5	A11	B8
High School 123	-8.8	0.7	A11	B8
High School 124	-7.6	-9.2	A11	B8
High School 125	-7.3	-9.1	A11	B8
High School 126	-7.1	-3.0	A11	B8
High School 127	-6.7	-1.5	A11	B8
High School 128	-6.7	-3.8	A11	B8
High School 129	-6.6	-5.1	A11	B8
High School 130	-6.5	-4.4	A11	B8
High School 131	-6.1	-4.3	A11	B8
High School 132	-5.9	-7.0	A11	B8
High School 133	-5.9	-1.0	A11	B8
High School 134	-5.8	-3.2	A11	B8
High School 135	-5.3	-4.7	A11	B8
High School 136	-3.9	-4.4	A11	B8

Participating Pennsylvania High Schools (continued)

High School Number	2010 PVAAS Math	2010 PVAAS Reading	2010 SIC A	2010 SIC B
High School 137	-3.8	-5.7	A11	B8
High School 138	-3.1	-1.5	A11	B8
High School 139	-2.9	-3.8	A11	B8
High School 140	-2.9	-2.8	A11	B8
High School 141	-2.5	-1.4	A11	B8
High School 142	-1.9	-10.5	A11	B8
High School 143	-1.9	1.3	A11	B8
High School 144	-1.5	-6.6	A11	B8
High School 145	-1.5	-5.4	A11	B8
High School 146	-1.2	-2.4	A11	B8
High School 147	-0.8	1.9	A11	B8
High School 148	-0.8	-4.0	A11	B8
High School 149	-0.5	-1.4	A11	B8
High School 150	-0.3	-0.5	A11	B8
High School 151	1.0	1.8	A11	B8
High School 152	1.8	1.3	A11	B8
High School 153	3.0	7.1	A11	B8
High School 154	4.0	3.0	A11	B8
High School 155	4.1	-3.5	A11	B8
High School 156	4.2	2.1	A11	B8
High School 157	4.3	0.6	A11	B8
High School 158	6.3	2.3	A11	B8
High School 159	6.6	-0.1	A11	B8
High School 160	9.7	1.1	A11	B8
High School 161	-10.6	-3.2	A2	B2
High School 162	-10.0	-6.6	A2	B2
High School 163	-9.5	-4.4	A2	B2
High School 164	-6.6	-3.6	A2	B2
High School 165	-6.1	-4.2	A2	B2
High School 166	-5.7	-5.4	A2	B2
High School 167	-5.6	-1.7	A2	B2
High School 168	-5.2	-3.0	A2	B2
High School 169	-5.1	-3.3	A2	B2
High School 170	-5.1	-7.2	A2	B2
High School 171	-5.0	-4.6	A2	B2
High School 172	-4.3	-3.1	A2	B2
High School 173	-4.0	0.4	A2	B2
High School 174	-3.7	-3.3	A2	B2
High School 175	-3.2	-4.2	A2	B2
High School 176	-3.0	-1.3	A2	B2
High School 177	-2.7	-1.5	A2	B2
High School 178	-2.5	-2.1	A2	B2
High School 179	-2.5	-1.7	A2	B2
High School 180	-2.4	1.1	A2	B2
High School 181	-2.3	-0.9	A2	B2
High School 182	-2.2	-3.3	A2	B2
High School 183	-2.2	-2.5	A2	B2

Participating Pennsylvania High Schools (continued)

High School Number	2010 PVAAS Math	2010 PVAAS Reading	2010 SIC A	2010 SIC B
High School 184	-2.1	-0.8	A2	B2
High School 185	-2.1	-1.7	A2	B2
High School 186	-1.9	-2.8	A2	B2
High School 187	-1.7	-1.8	A2	B2
High School 188	-1.7	-2.2	A2	B2
High School 189	-1.6	0.1	A2	B2
High School 190	-1.6	-2.8	A2	B2
High School 191	-1.6	0.8	A2	B2
High School 192	-1.5	0.5	A2	B2
High School 193	-1.5	-1.4	A2	B2
High School 194	-1.5	1.2	A2	B2
High School 195	-1.4	-2.2	A2	B2
High School 196	-1.1	4.1	A2	B2
High School 197	-0.9	2.3	A2	B2
High School 198	-0.7	-0.9	A2	B2
High School 199	-0.5	-1.7	A2	B2
High School 200	-0.5	0.4	A2	B2
High School 201	-0.5	1.4	A2	B2
High School 202	-0.4	0.3	A2	B2
High School 203	-0.3	0.6	A2	B2
High School 204	-0.3	-1.5	A2	B2
High School 205	-0.3	-1.0	A2	B2
High School 206	-0.2	-0.2	A2	B2
High School 207	-0.2	-8.0	A2	B2
High School 208	-0.2	-3.0	A2	B2
High School 209	-0.1	3.0	A2	B2
High School 210	-0.1	-0.7	A2	B2
High School 211	0.0	-4.6	A2	B2
High School 212	0.1	-1.8	A2	B2
High School 213	0.2	5.3	A2	B2
High School 214	0.3	2.4	A2	B2
High School 215	0.6	-0.6	A2	B2
High School 216	0.6	0.2	A2	B2
High School 217	0.7	1.7	A2	B2
High School 218	0.7	-0.4	A2	B2
High School 219	0.7	0.0	A2	B2
High School 220	0.8	0.9	A2	B2
High School 221	0.9	-1.7	A2	B2
High School 222	0.9	1.9	A2	B2
High School 223	1.1	0.9	A2	B2
High School 224	1.2	-1.4	A2	B2
High School 225	1.2	2.3	A2	B2
High School 226	1.2	-0.6	A2	B2
High School 227	1.3	-1.9	A2	B2
High School 228	1.4	-0.8	A2	B2
High School 229	1.4	0.7	A2	B2
High School 230	1.5	0.2	A2	B2

Participating Pennsylvania High Schools (continued)

High School Number	2010 PVAAS Math	2010 PVAAS Reading	2010 SIC A	2010 SIC B
High School 231	1.6	3.3	A2	B2
High School 232	1.6	6.5	A2	B2
High School 233	1.6	1.5	A2	B2
High School 234	1.6	2.5	A2	B2
High School 235	1.8	0.8	A2	B2
High School 236	1.8	-2.2	A2	B2
High School 237	1.9	-1.0	A2	B2
High School 238	2.0	4.0	A2	B2
High School 239	2.1	4.7	A2	B2
High School 240	2.1	1.3	A2	B2
High School 241	2.1	4.0	A2	B2
High School 242	2.3	1.4	A2	B2
High School 243	2.3	0.7	A2	B2
High School 244	2.3	0.5	A2	B2
High School 245	2.3	1.1	A2	B2
High School 246	2.4	-0.3	A2	B2
High School 247	2.6	2.5	A2	B2
High School 248	2.7	4.6	A2	B2
High School 249	2.7	-0.1	A2	B2
High School 250	2.9	0.6	A2	B2
High School 251	3.0	1.3	A2	B2
High School 252	3.1	-3.8	A2	B2
High School 253	3.2	1.7	A2	B2
High School 254	3.2	0.4	A2	B2
High School 255	3.4	0.3	A2	B2
High School 256	3.4	3.6	A2	B2
High School 257	4.1	2.4	A2	B2
High School 258	4.2	3.4	A2	B2
High School 259	4.2	1.6	A2	B2
High School 260	4.4	-2.8	A2	B2
High School 261	4.8	0.6	A2	B2
High School 262	4.9	1.7	A2	B2
High School 263	4.9	1.5	A2	B2
High School 264	5.0	0.5	A2	B2
High School 265	5.2	1.8	A2	B2
High School 266	5.5	2.2	A2	B2
High School 267	5.6	4.1	A2	B2
High School 268	5.8	5.5	A2	B2
High School 269	6.0	0.3	A2	B2
High School 270	6.2	1.0	A2	B2
High School 271	6.2	0.6	A2	B2
High School 272	7.0	4.6	A2	B2
High School 273	7.3	5.9	A2	B2
High School 274	7.5	5.1	A2	B2
High School 275	7.7	5.3	A2	B2
High School 276	7.7	5.3	A2	B2
High School 277	7.9	6.3	A2	B2

Participating Pennsylvania High Schools (continued)

High School Number	2010 PVAAS Math	2010 PVAAS Reading	2010 SIC A	2010 SIC B
High School 278	8.0	2.4	A2	B2
High School 279	8.2	4.7	A2	B2
High School 280	8.5	3.2	A2	B2
High School 281	9.0	4.6	A2	B2
High School 282	9.5	7.5	A2	B2
High School 283	10.1	9.4	A2	B2
High School 284	11.2	5.7	A2	B2
High School 285	16.4	6.0	A2	B2
High School 286	-7.1	-2.7	A3	B3
High School 287	-4.3	-1.2	A3	B3
High School 288	-3.9	-4.0	A3	B3
High School 289	-2.4	-2.2	A3	B3
High School 290	-2.0	15.0	A3	B3
High School 291	-1.7	-4.4	A3	B3
High School 292	-1.5	-1.9	A3	B3
High School 293	-1.2	-3.5	A3	B3
High School 294	-0.9	15.0	A3	B3
High School 295	-0.6	3.4	A3	B3
High School 296	-0.1	-0.5	A3	B3
High School 297	-0.1	1.2	A3	B3
High School 298	0.4	15.0	A3	B3
High School 299	0.7	-0.7	A3	B3
High School 300	1.0	-0.6	A3	B3
High School 301	1.5	3.1	A3	B3
High School 302	1.6	0.0	A3	B3
High School 303	1.6	-2.8	A3	B3
High School 304	1.8	-5.0	A3	B3
High School 305	1.8	-3.1	A3	B3
High School 306	2.4	1.3	A3	B3
High School 307	2.7	1.3	A3	B3
High School 308	3.2	15.0	A3	B3
High School 309	3.6	3.6	A3	B3
High School 310	4.2	3.1	A3	B3
High School 311	4.4	-0.4	A3	B3
High School 312	4.5	1.6	A3	B3
High School 313	6.1	4.5	A3	B3
High School 314	6.3	2.3	A3	B3
High School 315	6.5	6.5	A3	B3
High School 316	6.7	-1.0	A3	B3
High School 317	7.5	15.0	A3	B3
High School 318	8.0	15.0	A3	B3
High School 319	8.1	12.4	A3	B3
High School 320	8.7	6.1	A3	B3
High School 321	8.8	7.5	A3	B3
High School 322	8.8	15.0	A3	B3
High School 323	11.9	5.9	A3	B3
High School 324	12.0	2.0	A3	B3

Participating Pennsylvania High Schools (continued)

High School Number	2010 PVAAS Math	2010 PVAAS Reading	2010 SIC A	2010 SIC B
High School 325	-7.0	-4.3	A4	B4
High School 326	-6.4	-8.1	A4	B4
High School 327	-5.6	0.0	A4	B4
High School 328	-5.4	-5.4	A4	B4
High School 329	-5.3	-3.4	A4	B4
High School 330	-5.3	-3.5	A4	B4
High School 331	-4.9	-5.8	A4	B4
High School 332	-4.1	-1.9	A4	B4
High School 333	-3.9	0.1	A4	B4
High School 334	-3.7	-8.6	A4	B4
High School 335	-3.3	-3.0	A4	B4
High School 336	-3.3	0.9	A4	B4
High School 337	-3.2	-1.4	A4	B4
High School 338	-3.2	-0.1	A4	B4
High School 339	-3.0	1.3	A4	B4
High School 340	-3.0	-1.2	A4	B4
High School 341	-2.8	8.6	A4	B4
High School 342	-2.7	-0.2	A4	B4
High School 343	-2.5	0.5	A4	B4
High School 344	-2.5	0.8	A4	B4
High School 345	-2.4	-2.1	A4	B4
High School 346	-2.0	-3.7	A4	B4
High School 347	-1.9	1.9	A4	B4
High School 348	-1.5	-0.3	A4	B4
High School 349	-1.3	-1.7	A4	B4
High School 350	-0.4	0.5	A4	B4
High School 351	-0.3	-0.9	A4	B4
High School 352	-0.2	2.8	A4	B4
High School 353	0.7	-1.9	A4	B4
High School 354	0.8	1.4	A4	B4
High School 355	0.9	2.5	A4	B4
High School 356	0.9	-0.6	A4	B4
High School 357	1.1	-0.5	A4	B4
High School 358	1.3	3.2	A4	B4
High School 359	1.6	2.2	A4	B4
High School 360	1.7	1.7	A4	B4
High School 361	1.8	4.3	A4	B4
High School 362	1.9	1.2	A4	B4
High School 363	2.5	-0.8	A4	B4
High School 364	2.8	2.6	A4	B4
High School 365	3.3	-0.2	A4	B4
High School 366	3.3	0.6	A4	B4
High School 367	4.1	-2.0	A4	B4
High School 368	4.2	2.6	A4	B4
High School 369	5.1	-0.3	A4	B4
High School 370	5.4	-5.1	A4	B4
High School 371	6.0	9.4	A4	B4

Participating Pennsylvania High Schools (continued)

High School Number	2010 PVAAS Math	2010 PVAAS Reading	2010 SIC A	2010 SIC B
High School 372	9.1	-1.5	A4	B4
High School 373	20.0	3.0	A4	B4
High School 374	-7.1	-9.4	A5	B5
High School 375	-7.0	-3.2	A5	B5
High School 376	-3.9	-1.6	A5	B5
High School 377	-3.8	1.1	A5	B5
High School 378	-3.5	-3.3	A5	B5
High School 379	-2.7	4.4	A5	B5
High School 380	-0.5	-1.6	A5	B5
High School 381	0.1	0.6	A5	B5
High School 382	0.4	0.1	A5	B5
High School 383	1.1	-0.9	A5	B5
High School 384	1.4	-0.7	A5	B5
High School 385	1.9	-4.7	A5	B5
High School 386	2.0	-1.4	A5	B5
High School 387	2.3	4.8	A5	B5
High School 388	4.7	4.3	A5	B5
High School 389	4.9	4.4	A5	B5
High School 390	5.5	2.2	A5	B5
High School 391	6.6	2.5	A5	B5
High School 392	-6.8	-5.7	A6	B5
High School 393	-6.6	-6.3	A6	B5
High School 394	-5.6	-3.9	A6	B5
High School 395	-4.0	-0.8	A6	B5
High School 396	-3.9	-6.9	A6	B5
High School 397	-2.6	6.0	A6	B5
High School 398	-2.2	-1.2	A6	B5
High School 399	-1.5	-1.8	A6	B5
High School 400	-1.0	0.0	A6	B5
High School 401	-0.8	-2.1	A6	B5
High School 402	0.9	-0.6	A6	B5
High School 403	1.2	0.7	A6	B5
High School 404	4.1	6.0	A6	B5
High School 405	4.3	3.5	A6	B5
High School 406	5.7	4.7	A6	B5
High School 407	7.1	-4.5	A6	B5
High School 408	-5.8	-1.0	A7	B6
High School 409	-2.6	0.1	A7	B6
High School 410	-1.9	-2.1	A7	B6
High School 411	-0.9	0.7	A7	B6
High School 412	0.7	1.9	A7	B6
High School 413	4.2	7.0	A7	B6
High School 414	4.7	6.7	A7	B6
High School 415	5.8	6.9	A7	B6
High School 416	-5.5	-0.8	A8	B6
High School 417	-2.7	2.2	A8	B6
High School 418	-1.1	-4.0	A8	B6

Participating Pennsylvania High Schools (continued)

High School Number	2010 PVAAS Math	2010 PVAAS Reading	2010 SIC A	2010 SIC B
High School 419	-0.4	-1.2	A8	B6
High School 420	3.5	3.6	A8	B6
High School 421	13.1	1.9	A8	B6
High School 422	-5.8	-2.4	A9	B7
High School 423	-2.3	-1.7	A9	B7
High School 424	-2.0	1.6	A9	B7
High School 425	8.9	0.0	A9	B7
High School 426	11.6	5.2	A9	B7

VITA

Todd Matthew Davies

1564 Briarwood Lane
Pottstown, PA 19464
610-327-1001
tdavies@pgsd.org

PROFESSIONAL PROFILE:

A K-12 central office administrator with expertise as follows: district planning, curriculum and staff development, grade-level reconfiguration, school improvement planning, induction, mentoring, teacher supervision, public relations, and federal programs.

EXPERIENCE:

2011- Present: Director of Education & Assessment, K-12
Pottsgrove School District
- Implementation and Revision of Curriculum, Instruction & Assessment K-12
- Induction
- Federal Programs
- School Improvement Planning
- Common Formative Assessments
- Teacher Observation

2009-2011: Supervisor of Elementary Education, K-5
Pottsgrove School District
- Federal Programs
- Response to Instruction And Intervention
- Staff Development, Curriculum & Assessment, K-5
- School Improvement Planning

2007-2009: Coordinator of Federal Programs & Assessment, K-12
Pottsgrove School District
- Title I, Title II, & Title III
- School Improvement Planning
- Assessment and Data, K-12

2006-2007:	Staff Developer & Curriculum Assistant, K-12 *Pottsgrove School District* • Staff Development K-12 • Curriculum K-12 • Title I, Title II, & Title III • School Improvement Planning • Assessment and Data, K-12
2005-2006:	Fifth Grade Teacher *West Pottsgrove Elementary, Pottsgrove School District*
2002-2005:	Instructional Support Teacher *West Pottsgrove Elementary, Pottsgrove School District*
1996-2002:	Special Education Teacher *West Pottsgrove Elementary, Pottsgrove School District*

EDUCATION:

Doctor of Education:	Lehigh University	Spring 2012
Superintendent's Letter:	Lehigh University	Spring 2010
K-12 Principal Certification:	Saint Joseph's University	Summer 2006
Master of Education in Curriculum and Instruction:	Pennsylvania State University	Spring 2002
Bachelor of Science in Elementary/Special Education:	Kutztown University	Spring 1996

PROFESSIONAL ORGANIZATIONS:

- Association for Supervision and Curriculum Development (ASCD)
- Montgomery County Intermediate Unit #23 Instructional Council
- Montgomery County Intermediate Unit #23 Early Literacy Council
- Montgomery County Intermediate Unit #23 Math Council
- Montgomery County Intermediate Unit #23 ESL Council
- Montgomery County Intermediate Unit #23 Language Arts Council
- Pennsylvania Association of Federal Program Coordinators (PAFPC)
- Montgomery County Principals and Supervisors Association (MCPSA)