

RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Resources, Pupil-Type, or Personal Attention: Investigating the Relationship between  
School Size and Student Achievement on Pennsylvania Standardized Tests

by

Michael A. Ross

Submitted in Partial Fulfillment of the Requirements

For the Degree of

Doctor of Education

in the

Educational Leadership Program

YOUNGSTOWN STATE UNIVERSITY

May 2019

RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Resources, Pupil-Type, or Personal Attention: Investigating the Relationship between School Size and Student Achievement on Pennsylvania Standardized Tests

Michael A. Ross

I hereby release this dissertation to the public. I understand that this dissertation will be made available from the OhioLINK ETD Center and the Maag Library Circulation Desk for public access. I also authorize the University or other individuals to make copies of this thesis as needed for scholarly research.

Signature:

---

Michael A. Ross, Student Date

Approvals:

---

Dr. Karen Larwin, Thesis Advisor Date

---

Dr. Patrick Spearman, Committee Member Date

---

Dr. Daniel Keown, Committee Member Date

---

Dr. Jason Hilton, Committee Member Date

---

Dr. Sal Sanders, Dean of Graduate Studies Date

RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

©

M. A. Ross

2019

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

### ABSTRACT

This study examines whether a relationship exists between school size and student achievement utilizing the Pennsylvania System of School Assessment (PSSA) exams for students in grades 3-8 in both Reading and Math. More specifically, the study investigates the relationship of school size and achievement with historically underperforming subgroups, such as students who are economically disadvantaged, have a disability, or belong to a minority category. The study finds that students with disabilities had significantly higher achievement in smaller schools than larger schools. In fact, students with disabilities are as much as 30% more likely to score proficient on standardized exams in smaller schools than larger. This study also indicates that economically disadvantaged students are as much as 20% more likely to score proficient on standardized exams in smaller schools than larger. On the other hand, the study shows that students belonging to a minority category are more likely to score proficient on standardized exams in the largest schools compared to smaller schools. For example, Black students can be 20% more likely to score proficient in larger schools than smaller schools, and Hispanic students are 15% more likely. Although a correlation appears to exist between school size and student achievement, many confounding factors could exist that may affect the relationship and limit causal conclusions, such as school resources, class size, school climate, and teacher attitudes.

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

### ACKNOWLEDGEMENTS

Firstly, I would like to express my deepest gratitude to my committee chair, Dr. Karen Larwin. Thank you for your unwavering support and encouragement throughout this entire process. I'm inspired by your unparalleled dedication, motivation, and immense knowledge; I couldn't have done this without you.

I would also like to recognize the rest of my dissertation committee: Dr. Patrick Spearman, Dr. Daniel Keown, and Dr. Jason Hilton. I appreciate your insightful comments and poignant feedback, and most of all thank you for your unique perspectives. I will be forever grateful for your time and efforts.

I cannot begin to express my sincerest thanks to my family for their support, understanding, and faith. To my wife, Kristie, you have taken this journey with me (several times) and supported me every step of the way as I pursued my passions and ambitions. Thank you for believing in me, standing by my side, and sacrificing for me. To my wonderful daughters, Allison and Emily, although you may not have realized it, you supported and encouraged me to accomplish this goal. You inspire and energize me every day. I hope I showed you that hard work, dedication, and sacrifice are the keys to accomplishing your goals. To my parents, Joseph and Christine, thank you for supporting me in this and all my endeavors. I've always considered you my role models and mentors. You instilled in me the value of hard work, tenacity, and grit.

I would like to acknowledge the Union Area School District, its faculty, staff, and community. As a student, thank you for establishing in me the love of learning and the never-ending pursuit of bettering myself. As an employee, thank you for having confidence in my abilities and allowing me the opportunity to contribute.

TABLE OF CONTENTS

CHAPTER

1.

Introduction.....	1
Problem Statement.....	2
Purpose.....	3
Significance of the Study.....	4
Research Questions.....	4
Definition of Key Terms.....	5
Limitations.....	5

2.

Introduction of Literature Review.....	7
School Size History.....	9
Major Studies.....	12
A New Take.....	14
School Climate.....	16
Poverty.....	17
Learning Disabilities.....	20
Minority Students.....	22
Class Size.....	29
PA System of School Assessment (PSSA).....	32
PSSA Reliability and Validity.....	35

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

	PA Value Added Assessment System (PVAAS).....	42
	Chapter Summary .....	51
3.		
	Methodology.....	52
	Research Questions.....	52
	Participant Data.....	53
	Data Sources .....	53
	Data Collection Procedures.....	54
	Method of Analysis.....	54
	Limitations .....	55
	Chapter Summary .....	56
4.		
	Analysis of Data.....	57
5.		
	Discussion and Conclusions .....	75
	Summary of Findings.....	75
	Context of Findings.....	79
	Implications of Findings .....	83
	Limitations .....	84
	Future Research .....	86
	Conclusions.....	89

RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

REFERENCES .....	90
APPENDIX A.....	99
APPENDIX B.....	101



## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

### LIST OF TABLES

TABLE 1: Example School Level PSSA Data by Subgroup and Enrollment.....	53
TABLE 2: Overall Effect Size on School Size.....	58
TABLE 3: Overall Effect Size by Grade .....	59
TABLE 4: Estimated Effect Size Measure Across Reading and Math .....	60
TABLE 5: Estimated Effect Size Measure by Subgroup Factor .....	60
TABLE 6: Estimated Effect Size Measure on Gender by School Size .....	61
TABLE 7: Estimated Effect Size Measure on School Size by Gender and Subject.....	62
TABLE 8: Estimated Effect Size Measure on IEP and School Size .....	63
TABLE 9: Estimated Effect Size Measure on IEP by School Size and Subject .....	64
TABLE 10: Estimated Effect Size Measure on SES by School Size .....	65
TABLE 11: Estimated Effect Size Measure on SES by School Size and Subject .....	66
TABLE 12: Estimated Effect Size Measure on Race Overall by School Size .....	67
TABLE 13: Estimated Effect Size Measure on Race Specificity by School Size.....	68
TABLE 14: Estimated Effect Size Measure on Race by School Size and Reading.....	69
TABLE 15: Estimated Effect Size Measure on Race by School Size and Math.....	71

Chapter 1

School size and student achievement may be two of the most discussed educational topics in the past three decades. Educational experts, social science researchers, and politicians have debated the effectiveness of small schools and their relationship to common metrics like standardized test scores, attendance, and graduation rates; however, they also argue about the efficiency, fiscal prudence, and course offerings related to school size. For instance, in Iowa, the governor and over a dozen legislators endorsed proposals calling for minimum school district sizes, citing a need for improved student achievement and financial efficiencies (Johnson, 2006). On the other hand, advocacy groups such as the Rural School and Community Trust (2002) strongly contended that small schools have a positive influence on student achievement and minimal money is saved by consolidations.

A litany of research exists that addresses this vast subject, some of which addresses school size, overall, and others that address class size, specifically. Some commentate on the racial and socioeconomic achievement gap, urban versus rural schools, teacher attitudes and beliefs, and generalizability across states. One can find similarities in approach, such as the use of hierarchical linear modeling, regression, and correlation analysis. Although common threads exist between the research, much of the research concludes vastly different and contradictory results.

Gershenson and Langbein (2015) claimed their primary analysis provided no evidence that a relationship exists between school size and student achievement. Similarly, Scheerens, Hendriks, and Luyten (2014) claimed their study overwhelmingly showed that school size does not matter much for student achievement and other social

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

outcomes. On the other hand, Jones and Ezeife (2011) claimed a general decline in achievement can be seen as school enrollments increase for both inner-city and suburban schools. Some studies separately consider primary school achievement (Gershenson & Langbein, 2015), and others strictly consider high school size and achievement (Lee & Smith, 1996). Some only consider specific grade-spans and subjects like middle school science (Mann, Maxwell, & Holland, 2013). However, very little research exists that comprehensively evaluates primary, middle, and high schools and several achievement areas. Also, many research studies fail to control for subgroup variables such as SES, minority status, and students with individual education plans (IEP).

Additionally, much of the research has been conducted in several states including Iowa, Texas, North Carolina, and Washington, but few studies have been designed for Pennsylvania (PA) schools using PA's standardized tests. Additionally, no known research was discovered regarding PA School Performance Profile rating (SPP). This study will evaluate the association between school size and student achievement utilizing data specific to the state of PA including PA System of School Assessment (PSSA). In an era of increasing accountability and shrinking school budgets, school size must be considered.

### **Statement of the Problem**

As schools face mounting accountability regulations and shrinking school funding, leaders are looking to find environments that are most conducive to student achievement, and school size may be an option. Considering the conflicting views regarding the fiscal efficiency of consolidating schools and divisive research regarding the effect of school size on student achievement, educators are in need of more

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

comprehensive and overarching information. Some researchers contend that no relationship exists between school size and student achievement (Gershenson & Langbein, 2015), while others strongly assert that larger schools show lower achievement, larger achievement gaps, and lower graduation rates (Howley, Johnson, & Petrie, 2011). Many studies only consider particular types of schools, such as elementary or high schools, while others utilize narrow student achievement data, such as 8<sup>th</sup>-grade science or primary reading scores. Although a slew of research exists on school size, it has been constricted to just a few states, which questions the generalizability to the unique characteristics of PA and other states of similar demographics. A detailed description of the literature and any related school-size theory will follow in Chapter 2.

The previous research has demonstrated disparate results and limited achievement variables, but this study will utilize a comprehensive approach that will compile data from varying school types including elementary, middle, and middle/high schools. The study will also look at wider-ranging standardized achievement results including English language arts, and math. The study will also investigate how school size affects other influential variables such as ethnicity, socioeconomic status (SES), and students with disabilities.

### **Purpose**

The purpose of this study is to evaluate the impact of school size on student achievement using PA standardized tests in three types of schools including elementary, middle, and middle/high schools. Data will be compiled from PSSA exams (grades 3-8) in the subjects of math, and English language arts. The study will utilize correlation

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

analysis to evaluate subgroups of students including socioeconomic status, race, and special education status.

### **Significance**

From a methodology standpoint, this study is structured similarly to other studies (linear modeling, correlation, and regression), but the study is a direct shift from typical achievement data. Other notable researchers have utilized stock, national achievement data (American College Testing [ACT] and Scholastic Aptitude Test [SAT]) and state-mandated testing data, but no known research has utilized PSSA data. Finally, almost a quarter of PA's public-school districts have less than 1,200 students. Evaluating their effectiveness versus the 25% of districts that have more than 4,000 students is a common policy and legislative debate in PA.

### **Research Questions**

The research questions for this study are:

1. Do students who receive free/reduced lunch demonstrate significantly higher student achievement on standardized tests in smaller schools compared to larger schools?
2. Do students who have an IEP demonstrate significantly higher student achievement on standardized tests in smaller schools compared to larger schools?
3. Do students who belong to a minority racial status demonstrate significantly higher student achievement on standardized tests in smaller schools compared to larger schools?

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

### **Key Word Definitions**

*Achievement Gap* - the difference in standardized test scores between historically underperforming students and other students (Pennsylvania Department of Education [PDE], 2017)

*Keystone Exams* - end-of-courses' exams for PA students in algebra 1, literature, and biology (PDE, 2017)

*Low Socioeconomic Status (SES)* - students who qualify in PA for free/reduced lunch (PDE, 2017)

*PA System of School Assessment (PSSA)* - annual, standards-based, criterion-referenced math and English language arts exams given to all PA students, in grades three through eight, and, in science, in grades four and eight (PDE, 2017)

### **Limitations**

Although a need exists to evaluate PA schools, the external validity of this study would be suspect outside of the sample. Using the free- and reduced-lunch rates is a fairly vast metric, and it does not necessarily offer potentially valuable details. For example, families may qualify for free and reduced lunch with annual salaries of 100% up to 185% above the poverty line. Differentiating the differences between the poorest students and those who barely qualify may be useful, but, impossible with this structure. The simplistic correlation and regression analysis does not allow for any school culture analysis or gather any qualitative data. As with any linear analysis, the PSSA scores would be sensitive to outliers.

Although the data may suggest a relationship exists between school size and student achievement, merely identifying the size of the school as the reason for higher or

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

lower achievement would be premature and overly simplistic. As the literature review will suggest, a myriad of factors have an influence on student achievement and many factors, including school size, may create the existence of those factors and the influence they have on student achievement. Issues including but not limited to school climate, teacher attitude and beliefs, inclusionary methods, and others may have a substantial impact on student achievement; however, the size of the school may increase/decrease the likelihood those and other factors exist. Considering this dynamic, the study may illuminate trends in regards to the relationship of school size and student achievement and may highlight the potential existence of external factors effecting achievement, but the scope of the study will not permit assertions on external factors nor will it allow for the elaboration on those factors.

## Chapter 2

### **Review of Literature**

The review of literature will examine the relationship between school size and student achievement and will also address theoretical constructs such as SES, special education status, minority student performance, and performance achievement gaps. The review will also evaluate the effect of school size on school culture, as well as gauge the association of class size.

A plethora of factors exist that historically hinder student achievement and widen the academic achievement gap. For instance, low-income students perform less well on academic measures such as standardized test scores, grades, graduation rates, and college enrollment than their high-income counterparts. In fact, the achievement gap in regards to socioeconomic status has widened over the past 30 years (Reardon, 2013). In evaluating minority performance on academic measures, Reid (2002) found that Black students were 52% proficient on standardized tests compared to 82% of White students and scored 200 points less than White students on SAT exams. Additionally, Black students had an exponentially higher drop-out rate, accounting for nearly 35% of all dropouts, although accounting for only 31% of the student population. Finally, Black students were found to be drastically underrepresented in advanced/honor courses than White students. In fact, 14% of course enrollments were comprised of Black students compared to 85% of White students. Moreover, students with IEPs also showed a major disparity in student achievement. For example, students with IEPs have a 13%



## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

proficiency rate compared to students without IEPs, which demonstrates a proficiency rate over 50% (Riggen, 2013).

In a 2017 study by Tienken et al., these researchers were convinced that standardized tests were so strongly influenced by non-school factors that their results were not objective in the least. Using a correlational, explanatory, longitudinal study, the researchers were able to accurately predict the scores of middle schools simply based on narrow pieces of information regarding school demographics. The study used only three demographics contained in the U.S. Census, including percentage of families with an income over \$200,000 a year, percentage of people in the community in poverty, and percentage of people in a community with bachelor's degrees. Tienken et al. claimed the use of standardized test results is an overly simplistic, mechanistic view of a complex and intricate educational environment (Tienken et al., 2017). Although many social, emotional, physiological, physical, and ecological influences are connected to educating a child, much of educational policy is constructed solely on the premise of test results.

White et al. (2016) argued the impact of immutable factors such as SES or race have a major impact on test scores. Exploring the test scores from 452 schools across the state of New Jersey, these researchers determined, by high school, 52% of variance in language arts, and nearly 59% of math exams were accounted for by SES and racial factors. Considering this impact, just a 1% increase in school minority population can correspond to a .19 decrease in proficiency in language arts and .33 for math. With implications this significant, the researchers suggested improving academic achievement requires a long-term approach and short-term plans and will likely be stymied by these outside factors (White et al.).

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Many school stakeholders are investigating ways to counteract these effects. Lawmakers are thrust into debates ranging from school funding to charter schools, and school officials are developing vast interventions to make improvements. Several research studies have concluded that smaller schools may not only better serve all students, but may, in fact, help alleviate achievement gaps with students of color, those in poverty, and those with disabilities. As the school-size debate raged on over the past century and will continue to do so, the answer for combatting achievement gaps may lie in the size of the school.

### **School-Size History**

At the turn of the 20<sup>th</sup> Century, a push for bigger public schools began and the number of school districts began to drastically drop. In 1932, nearly 130,000 school districts existed; that number fell below 20,000 by the early 1970s and now stands under 14,000. By and large, the push toward larger districts was driven by a desire for improved efficiency, but, by the 1980s, the nation's public schools were spending, when adjusted for inflation, nearly 10-times per pupil than in the 1920s. More recently, researchers have determined that, although some economies of scale exist for larger schools, only modest fiscal savings are apparent for very small schools (Coulson, 2007). As the per-pupil cost continues to rise and school consolidations are considered by legislators, researchers have turned their attention to the viability and potential benefit of small schools over large schools. In fact, several studies conclude numerous advantages exist in smaller schools varying from improved student engagement to a lessened disparity between historically underperforming students and their peers.

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

PA lawmakers followed the national trend of school consolidations; the most major shift in school consolidations occurred in the 1960s with the passing of consolidation laws in 1961, and, again, in 1963 (Joint State Government Commission, 2017). Some lawmakers argued for decentralized, local control of smaller schools, while others asserted that larger schools could offer a more comprehensive, and more affordable education. Considering the historical context of a post-World War II PA, lawmakers ultimately made the shift toward larger schools. With the intention of making more college-ready graduates and preparing students for a global technological battle with Soviet Union, the PA legislature opted to create larger schools to meet their demands and do so in a more fiscally responsible manner (Joint State Government Commission, 2017).

After what historians referred to as contentious debate, Act 561 of 1961 mandated school district consolidation across the commonwealth. The student minimum was established at 4,000 students with the potential to lower the minimum to 2,500 if certain topographical, student, socioeconomic, or transportation criteria existed. It was considered a conservative approach, new school districts could only develop from consolidation not a massive geographical redistricting. Ultimately, a county board of school directors had responsibility for determining the overhauled districts (Joint State Government Commission, 2017).

During the development, and, for the years following the consolidation requirements, strong opposition was displayed from various stakeholders. In 1963, some major concerns of the 1961 language were addressed in a 1963 revamping of the law. Act 299 made the 4,000-student limit more of a recommendation rather than requirement,

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

and the law developed a process for aggrieved districts to appeal merger decisions (Joint State Government Commission, 2017).

Following the new laws, PA saw a reduction in school districts from 2,277 in 1960, to 669 by 1970. By the end of the 1970s, that number was reduced to 505, and, because of a federal desegregation lawsuit, the number of districts was dropped to 501. In the only voluntary district merger in the state's history, the Center and Monaca School Districts merged in 2009 to form the Central Valley School District, which dropped the number of districts to 500, currently (Joint State Government Commission, 2017).

In the mid-to-late 2000s, PA Governor, Ed Rendell, suggested that school districts should be merged from the 500 to 100. In response, the legislature decided to study the situation and plausibility of another mandated consolidation effort. Reports from the Pennsylvania School Board Association ([PSBA], 2009) and Standard and Poor's School Evaluations Services (2008) provided both mixed and unexpected results. According to the PSBA (2009), the 1960s' consolidations lacked any empirical information regarding financial savings or educational achievement. PSBA (2009) framed in the consolidations, from a historical perspective, as losing a sense of community as schools became so large that parents felt detached and uninvolved.

Standard and Poor's (2007) completed an elaborate fiscal audit of PA districts to evaluate the potential cost-savings of district consolidations. Generally, the report identified 88 school districts with 97 different combinations where mergers might be desirable and/or possible. The researchers went on to say that district consolidations are highly individualized and contextual, and districts willing to consolidate should be investigated more specifically. However, the report indicated that optimal district size

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

appeared to be under 3,000 students. In fact, districts between 2,500 and 3,000 students had the lowest average cost per student. The largest and the smallest districts typically had the highest cost per student.

The Joint State Government Commission (2017) divided district sizes into five categories: very small, small, medium, large, and very large. Nearly half of PA schools fall under the medium category, or, between 1,500 and 3,000 students. Very small districts, (less than 500 students), account for 2.4% of PA districts, while small districts (500-1000 students) make up 13.2 percent of districts. Large districts, those with 3,000 to 10,000 students, account for nearly one-third of PA school districts, while very large districts (10,000+ students) make up just 3.4% of districts.

### **Major Studies**

In an historic study conducted by Howley and Bickel (2000), they analyzed 29 sets of test scores across four states and found that as schools become larger the negative effects of poverty on student achievement become more profound. Specifically, they indicated that the effects of poverty on student achievement could be 10-times stronger in larger schools than smaller schools. Howley and Howley (2004) set out to reexamine their 2000 study by using students as the unit of analysis. Their study drew similar conclusions claiming that SES is more likely mitigating in terms of student achievement in smaller schools, and determined these effects, although slightly less robust in rural schools, are consistent across school type. Contrary to previous conclusions, they noticed data that suggested larger schools may benefit students in the highest SES levels.

Abbot, Joireman, and Stroh (2002) attempted to replicate the Howley and Bickel (2000) study utilizing the same hierarchical, linear-modeling structure, but wanted to

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

evaluate the results for students in the state of Washington, in different grade spans. Similarly, the results signified that larger district size was detrimental to achievement with Washington 4<sup>th</sup>- and 7<sup>th</sup>-graders, and the negative relationship between SES and student achievement was strengthened in larger schools. However, Abbot et al. did not find that larger schools are more advantageous than smaller schools with affluent students.

Valerie Lee, with Smith (1996), and Loeb (2000), wrote extensively on the subject of school size and student achievement. In 1996, Lee and Smith attempted to identify the optimal school size and found significant evidence that regardless of social background, students demonstrate higher achievement in schools with 600-900 students. As with other studies, Lee and Smith (1996) found that poorer, minority students are worse served in larger schools with more than 2,100 students. Lee and Loeb (2000) offered some interesting insights on school size. Rather than evaluating differences in student achievement, the authors assessed the fluctuation of teacher attitudes and beliefs. Teachers in smaller schools were found to have a more positive attitude, were more apt to feel responsible for student learning, and provided more personal attention to students. Furthermore, Gershenson and Langbein (2015) claimed that smaller schools had higher graduation rates related to improved school culture and other non-cognitive skills such as motivation and grit. Although these authors were attempting to evaluate the effect of school size on student achievement, each found that smaller schools created characteristics in teachers and students that contributed to a positive school culture and climate.

### **A New Take**

Several studies have utilized a similar structure as the previous studies, but each has contributed their unique twist on the research. In a more recent study, with a slight adjustment accounting for grade-level enrollment, Gershenson and Langbein (2015) found no causal link between school size and overall student achievement but did find that two distinct subgroups are most

harmed by larger school size: economically disadvantaged students and students with learning disabilities. Specifically, the math and reading achievement of students with a learning disability and the reading achievement of economically disadvantaged students are disproportionately hurt by increases in school size and this effect does not vary between geographic location (Gershenson and Langbein, 2015, p. 151).

An Alspaugh and Gao (2003) study took a narrow data set by only using 5<sup>th</sup>-grade achievement in one Missouri school district, but the structure carried a unique reason. The researchers were attempting to evaluate 39 schools but controlled for other potential confounding variables because the district has similar administrative structures, identical curriculum, and resource allocation. The researchers found significant differences in achievement based on enrollment figures and similarly to other studies, noticeable differences related to SES.

Johnson (2006) attempted to dispel the myth that larger high schools, because they could offer more courses and more resources, could more effectively educate students. Johnson (2006) criticized other researchers who came to this conclusion by using national standardized tests such as the ACT. Johnson (2006) recognized that

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

larger, more affluent school districts offer more ACT prep courses and their students are more likely to afford retakes as reasons for higher ACT performance. When using state-mandated achievement tests, that nearly all students take, Johnson (2006) found no significant difference between larger and smaller high schools, hence finding evidence that vast course offerings and services had little effect on student achievement.

Uerling and Glugosh (1999) also tried to dispel increased-course offerings during a Nebraska consolidation initiative. On average, the researchers found that larger school districts offered a broader curriculum but had lower participation rates across the curriculum compared to smaller schools. In other words, larger schools have the resources to offer more classes, but they do not need students enrolled to still offer the classes. A vast curriculum appears to benefit a sliver of the student population but means less for most other students (Uerling & Blugosh, 1999). Johnson (2006) asserted that more course offerings can allow students to “hide” in a wide array of courses. She found that larger schools seem to homogeneously track students and offer classes for advanced or college bound students (p. 10). For instance, schools may offer high school English as a “college prep” class or as a “general” class for other students, but many bright students take the “general” class as a way of avoiding the more rigorous offering (p. 10).

Smaller schools and larger schools may have the same tendency to have more offerings and advanced courses (Iatarola, Conger, & Long, 2011). Iatarola et al. suggested the amount of course offerings is more contingent on the demand from students rather than the size of the school. Iatarola et al. found that demand is highly related to achievement; in turn, students who have performed well on standardized exams tend to seek out more advanced coursework. With this in mind, their analysis indicated



## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

that larger schools can actually offer less advanced coursework should demand drop.

When they found a relationship between school size and course offerings, they also found higher achievement in those schools.

Many studies utilized reading and math achievement as the dependent variable, but Mann et al. (2013) utilized 8<sup>th</sup>-grade science achievement in Texas to evaluate school-size effect. Using quantitative methods, the researchers determined a statistically significant difference existed between school size, and, in particular, the highest performing students were enrolled in medium-size school districts as opposed to large or mega school districts. In a 2016 study by Wood, Kiperman, Esch, Leroux, and Truscott, these authors investigated other potential negative effects of larger schools and found, when controlling for variables such as race and socioeconomic status, that larger schools were associated with higher dropout rates.

### **School Climate**

According to the National School Climate Council, four major factors shape school climate: safety, relationships, teaching/learning, and the institutional environment (Cohen, Pickeral, & McCloskey, 2009). Developing a positive and supportive school climate is crucial for student academic achievement, and students who feel safe, connected, and engaged are more likely to learn better. In fact, students demonstrated higher achievement, lower dropout rates, less behavioral concerns, higher attendance, and less reported instances of bullying (Cohen et al.). Although school leaders can take several steps to improve school climate, many researchers contend that achieving optimal levels of school climate are much more likely in smaller schools than larger schools. In fact, students are able to develop and maintain positive social relationships and are more

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

likely to consider themselves to be part of the community when they attend smaller schools (Abbott et al., 2002).

Using a mixed-methods design, Goldkind and Farmer (2013) studied parental perceptions of their children's schools using a learning environment survey. Parents in smaller schools believed the educational environment was safer, more respectful, and offered more opportunities for parental involvement. In more substantive terms, the authors indicated that schools characterized by less safety and respect create a school climate that suppresses the opportunities for communication and involvement. For example, smaller schools are able to emphasize relationships with their stakeholders, prioritize the school/parent partnership, and promote parental participation, which enhance academic achievement (Goldkind & Farmer, 2013). The authors continued to say that with an upsurge of policy-making focusing on the creation of smaller schools, it is important to understand the impact enrollment can have on school climate and, ultimately, student achievement. Other researchers have also concluded that school size and school climate are directly related to student achievement. Jones and Shindler (2003) so strongly supported the concept that they claimed school climate is one of the most predictive factors in any school's capacity to improve student performance.

### **Poverty**

Many researchers have concluded that student poverty has a significant relationship with student achievement, and, in a 2014 study conducted by Silvernail, Sloan, Paul, Johnson, and Stump, they stated the evidence clearly showed that as the percent of poverty increases, student performance decreases. They went on to say that poverty level is the single, best predictor of average student achievement, but other

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

factors may exist that help alleviate the effects of poverty. Web and Thomas (2015) asserted that three socioeconomic factors, which include poverty, nutrition, and self-esteem, have a notable impact on the achievement of Black males. Generally, poorer students lack adequate nutrition and medical care, as well as consistency and continuity in the home. The researchers continued to state that poor children often begin their educational careers with weaker language skills mostly accredited to minimal exposure to enhanced vocabulary and limited access to developmentally appropriate reading materials. Consequently, living in poverty is a significant, environmental stressor that has a detrimental effect on a child's emotional, academic, and social development.

Teaching students in poverty is a large-scale issue in the United States. Darling-Hammond (2015) estimated that nearly two-thirds of U.S. middle school teachers work in schools where at least 30% of the students are considered economically disadvantaged. This rate is one of the highest in the world and nearly triples the average from around the world. In comparison, the nearest rates to the United States are Malaysia's and Chile's. Darling-Hammond (2015) went on to make the point that these astronomical rates are largely ignored by law and policy makers. In fact, the rates of students below the poverty line (25%) and students experiencing homelessness are rapidly growing, which means more and more children do not have regular access to food or healthcare, and many are surrounded by violence and drug abuse. The challenge for educators is that they now spend a significant portion of their time helping children and their families navigate these issues but are still charged with closing achievement gaps (Darling-Hammond, 2015).

Many researchers have found that school size may act as a mitigating factor with economically disadvantaged students.

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

In a landmark study, Howley and Bickel (2000) analyzed tests scores from Georgia, Ohio, Montana, and Texas to further examine the relationship between school-level performance on tests and poverty level. The study found the negative effects of poverty increase as schools become larger, and, in fact, low achievement is as much as 10-times stronger in larger schools than in smaller schools, particularly in the middle-level grades. Howley and Howley (2004) set out to recreate their earlier studies, but wanted to use students as their unit of analysis. Similarly, the results suggested that smaller school size creates an achievement advantage for poorer students and helps mediate the powerful relationship between SES and achievement. In a substantial claim, Howley and Howley (2004) indicated that most of at least 12 state-level investigations into school size, student achievement, and SES showed effect sizes favoring smaller schools and smaller districts in poorer communities.

In another large state-level study, the Rural School and Community Trust (2002) investigated the possibility of school size as a potential mitigating influence regarding poverty. This research claimed four major findings of the study. Using regression analysis, higher levels of poverty in a school community were found to be more detrimental to student achievement as school enrollment grew. Data also showed the achievement gap between poorer students and more affluent students was narrowed in smaller schools/districts and widened in larger schools/districts. Although the research produced strong evidence that smaller schools are more effective at combatting the poverty effect, they found that smaller schools, in smaller districts, are the most effective. In other words, a smaller school in a large district does not equalize the effects of poverty on student achievement as much as a smaller school in a small district. Finally, the

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

researchers discovered data that suggested the relationship between school size, poverty, and student achievement is three-times greater in schools with the largest percentage of Black students (Rural School and Community Trust, 2002).

In other smaller scale studies, researchers found a strong influence of poverty over achievement, but also recognized the effect of the influence is dependent on the size of the district. The results suggested that increased school size increases the power of poverty over student achievement and widens the achievement gap between wealthier and poorer students (Johnson, 2006). In a study with more modest results, Coladarci (2006) found that smaller school size helped mitigate the poverty affect in mathematics standardized tests but did not show a relationship with reading scores. In a Council for Great City Schools' (2012) study, researchers found that students in poverty not only demonstrate lower achievement results by average, but they also experience other variables that can potentially dictate achievement results. In fact, students in the most impoverished, urban school districts have higher school enrollments, 93 more students on average (589 versus 486). Furthermore, they experience higher student-teacher ratio of 1:18 versus 1:16 national average (p. 26).

### **Learning Disabilities**

Closing the achievement gap between students with disabilities and students without disabilities has been of interest to educators for decades. In Illinois, for example, students with IEPs also showed a major disparity in student achievement. Students with IEPs have a 13% proficiency rate compared to students without IEPs who demonstrate a proficiency rate over 50% (Riggen, 2013). Furthermore, Kirby (2017) contended that students can experience achievement gains when schools implement inclusionary

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

techniques, but many school districts overwhelmingly utilize separate, self-contained classrooms.

Although many interventions exist to address the achievement gap, many researchers find that smaller schools are able to more effectively intervene with students having IEPs. Even in studies that do not find a causal relationship between school size and student achievement, researchers have found evidence that students with learning disabilities may suffer from increases in school size. The scores of students with learning disabilities in both reading and math are disproportionately harmed by increases in school size (Gershenson & Langbein, 2015). In fact, Gershenson and Langbein (2015) claimed:

that students with learning disabilities were particularly sensitive to increases in school size since larger schools are not as able to match exceptional students to appropriate support programs, or because students with disabilities are more susceptible to the weaker social bonds likely inherent in larger schools. More specifically, a 10-student increase in a grade was found to reduce the math and reading scores of exceptional students by a practically significant level (p. 151).

As Kirby (2017) contended, students with disabilities show higher achievement when schools utilized inclusionary instructional techniques, although most schools implement pullout procedures because of staffing and other logistical prohibitions. According to Marks, Kurth, and Bartz (2014), the most important feature influencing placement in inclusive or segregated settings is often district size. In fact, larger districts have lower percentages of IEP students spending a large percentage of their day in general education settings. Additionally, smaller schools tend to embrace philosophies

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

that encouraged inclusive education, such as focusing on the ability of students, having more consistencies in beliefs between teachers and the leadership team, and the presence of inclusive programs. Using Kirby's (2017) statistics, students with IEPs experienced educational gains when included in the general educational environment. For instance, elementary school students with learning disabilities, when included in general education, had higher reading scores than students in self-contained special education classrooms. This trend also continued with math scores with students in 8<sup>th</sup> grade.

### **Minority Students**

Just as students experiencing poverty demonstrate an achievement gap on standardized tests, minority students show similar gaps in achievement. Data suggest that many minority students continue to perform more poorly on standardized tests compared to their White counterparts, including exams like the SAT. Additionally, minority students continue to have disproportionately high rates of discipline infractions and high school drop-out instances, as well as an underrepresentation in honors' and advanced courses (Reid, 2002). Trends in mathematics achievement indicate an achievement gap exists between minority students and White students, and this evidence also indicates the gap may be widening (Webb & Thomas, 2015). The National Center for Educational Statistics (NCES) in 2015 showed that Black and Hispanic students continue to lag behind White students by more than 20 points on the National Assessment of Educational Progress Exams (NAEP) math and reading assessments in 4<sup>th</sup> and 8<sup>th</sup> grades (Bohrnstedt, Kitmitto, Ogut, Sherman, & Chan, 2015). The 20-point gap equates to approximately two grade levels. According to NAEP's analysis of the 2009 results, Black boys scored comparably as Black girls in the same grades, but they also showed a 30-point gap

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

between White boys in those grades, which equate to nearly three grade levels (Webb & Thomas, 2015).

In the contributory Student Teacher Achievement Ration (STAR) experiment from Tennessee, researchers utilized over 25 years of data to evaluate school size on student achievement (Achilles, 2012). Many researchers have reviewed and summarized the results of the STAR study, but Achilles (2012) made several assertions about minority student achievement. The author indicated that student achievement was significantly higher for students in smaller schools for all students but was particularly beneficial for minority students. In the short term, Achilles (2012) contended that minority students showed improved test outcomes, more school engagement, and reduced grade retention and dropout rates. In the long term, Achilles (2012) argued that students in smaller schools were more likely to take the SAT and ACT exams, and the benefit for Black students was substantially greater than for White students. Specifically, the Black-White gap in college entrance tests was reduced by 54%. The graduation rate for students attending smaller schools, for more than three years, nearly doubled for low-income minority students, essentially closing the achievement gap on this measure.

In a Howley and Howley (2004) study, they found that minority students in smaller schools showed higher achievement in both reading and math compared to minority students in larger schools. More specifically, mathematics scores for Black and Latino students were the highest in schools enrolling 600-1200 students, but achievement equity was maximized in schools enrolling 300-600 students. Schools larger than 2,100 students showed both the smallest gain and the greatest inequity. Reading scores showed similar and slightly different results. The highest gains were realized in schools



## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

enrolling 600-1200, the same as math scores; however, achievement equity was maximized in schools with less than 300 students, and both gains and equity were drastically reduced in schools larger than 1500 students (Howley & Howley, 2004).

In a study specifically designed to analyze Hispanic student achievement, Zoda, Slate, and Combs (2011) found results contrary to other similar studies. The researchers found statistically significant differences in achievement between Hispanic and White students, however, the Hispanic student achievement in reading and writing showed higher passing rates in large schools than in very small schools. Though not statistically significant, Hispanic student achievement in math had higher-than-average passing rates in large schools than in small schools (Zoda et al.). Contrary to previously reviewed studies, the researchers found conflicting results and none of the small schools had statistically higher achievement compared to large schools.

In a 2015 study conducted by the National Center for Educational Statistics, researchers investigated data from the NAEP Mathematics Grade-8 Assessments to look further into achievement of students as it related to the percentage of students in the school who were Black. This percentage, nationally, referred to as racial density, shows disparate differences for Black and White students. For instance, White students, on average, go to a school where 9% of the population is comprised of Black students. On the other hand, Black students attend schools where 48% of the student population is Black. Racial density in PA is even more contrasting. White students attend schools where only 6% of the population is made up of Black students, and Black students attend schools with 62% of the population being Black. NCES researchers referred to this concept as the re-segregation of American schools (Bohrstedt et al., 2015).

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

For the 2015 NCES study, researchers developed four density categories to examine schools nationally. Schools were designated in four percentage ranges regarding the percentage of Black students: 0-20, 20-40, 40-60, and 60-100. Nationally, over three quarters of U.S. schools fall into the lowest density category and 10% fall into the highest density category. However, the average percentage of students in a school in the highest density range varies by region. In the south for instance, 46% of schools have been determined to be high-density schools, but schools in the west only have 2% of schools in that range. Thirty-three percent of midwest schools were considered high-density schools followed by 19% for the northeast region.

Researchers examined NAEP Mathematics Grade-8 data to discover any potential relationship between Black students and achievement. Researchers decided to evaluate mathematics achievement based on the assumptions that mathematics exams are less influenced by English-language-learner (ELL) status. The findings indicated that both Black and White students in the highest-density schools demonstrated lower achievement than students in the lowest-density schools. Specifically, White and Black students in high-density schools scored on average 10 points less than students in low-density schools; however, the achievement gap between White and Black students was negligible between the lowest- and highest-density schools, in particular, 25 points versus 26 points, respectively.

As discussed above, many researchers have contended a relationship exists between student achievement and other variables such as SES. NCES researchers attempted to address the complexities of these associations by evaluating the effect of Black student density while controlling for SES. Within every density category, a higher

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

percentage of Black students were eligible for free and reduced lunches than White students. For example, in the highest-density schools, 83% of Black students were eligible compared to just 53% of their White counterparts. When controlling for SES, NCES researchers found that White student achievement remained relatively unchanged while Black student achievement was significantly higher across all density levels. Furthermore, the previously observed achievement gap between Black and White students narrowed by 5-7%, most significantly in the 20-40% category.

The NCES (2015) study also investigated varying degrees of characteristics that could have a more broad effect on achievement, including teacher/school characteristics. These variables include “teacher qualifications, instructional practices, resources, and school climate (NCES, 2015, p. 18).” While controlling for these variables, these researchers discovered several intriguing results. Achievement gaps were reduced in every density category by at least 6%, and, at most, by 10%. The largest achievement gaps were still the largest in the two highest-density categories but were significantly lower than gaps observed without controls. Although White student achievement did not change between the lowest- and highest-density schools, Black student achievement was lower in the highest-density categories compared to the lower density. While controlling for teacher/school characteristics, researchers also delineated results by gender. White females tended to score higher than Black females, but a significant relationship was neither established regarding Black student density and achievement, nor density levels and the achievement gap. However, with males, Black students’ achievement was not only lower, but the achievement gap between White and Black students significantly widened.

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Although the NAEP researchers were not attempting to offer suggestions how to close achievement gaps, they contended school administrators should understand the differences regarding “between-school” and “within-school” characteristics (p. 22). Between-school characteristics refer to disparities in critical resources such as technology, updated textbooks, or teacher qualifications. Within-school traits would involve differential teacher expectations, ability tracking, and school climate. Although conventional wisdom may suggest that most achievement gaps would be associated with between-school variables including resource allocation, the researchers indicated that nearly half of the average achievement gaps can be attributed to within-school differences. Administrators would be better served by focusing efforts more surgically on the specific types of differences as a method to narrowing achievement gaps.

Exploring teacher expectations has been a heavily cited phenomenon, especially considering that expectations could carry implicit or explicit biases. Peterson, Rubie-Davies, Osborne, and Sibly (2016) examined the effects of explicit ethnicity-based expectations for student achievement as well as implicit prejudiced attitudes. Utilizing both a traditional teacher-expectation measure and an implicit association task, researchers determined that students’ performance can be positively and negatively impacted by teacher attitudes and perceptions. For instance, students performed better in classrooms where teachers had high expectations and implicit biases favored the student, ethnic background. Conversely, students in classrooms with teachers who demonstrated a significant, implicit, prejudiced attitude toward ethnic stereotypes, student performance was predictable.

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

In a study by Bol and Berry (2005), they found students' test scores on a 6<sup>th</sup>-grade standardized mathematics test were reliant on teacher expectations and perceptions. The researchers went on to say that Black male students were victims of decreased opportunities because of the negative perceptions of their teacher. In fact, in addition to lower achievement, Black males received limited recommendations to advanced mathematics courses. Webb and Thomas (2015) suggested that teachers expect three-times more from White students than from Black students.

Other researchers have discussed the cultural and legislative shift to high-stakes standardized testing and accountability measures. Wasserberg and Rottman (2016) contended this heavy reliance on testing and teacher ratings has forced teachers to utilize classroom practices centered on test preparation. Classrooms adopting this test centric approach can lead to teacher underestimation of student academic ability, which, as stated before, can contribute to the underperformance of minority students. Wasserberg and Rottman (2016) continued to say teacher underestimation of student ability can, not only cause underperformance, but, can also create a dynamic where students underestimate their own ability. Jordan and Lovett (2007) referred to this dynamic as a stereotype threat and expounded by saying that negative stereotypes about one's group may lead to thoughts to conform to the stereotypes, hence, leading to the failure that is feared. Jordan and Lovett (2007) used the following example to help illustrate:

For example, a girl who tries out for an improvisational comedy troupe in high school, aware of a stereotype that women do not make good comedians, may find herself flustered and unable to perform to her potential. She may, furthermore,

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

attribute her comedic stumbles to her femaleness and become dejected in a way that male members of her troupe do not when they experience frustration. (p. 46)

Considering issues such as biases, underestimation, and more vast concepts such as school climate, school size may offer an inherent intervention to help mitigate the negative effects on those and similar dynamics.

### **Class Size**

Evaluating the effect of school size on student achievement cannot be completely separate from considering class size. Having a small school does not necessarily constitute having small-class sizes. For instance, a school with 600 students and 30 teachers has a class-size average of 20 students per teacher. On the other hand, a smaller school with just 300 students with 12 teachers creates student-teacher ratio of 25. Although the latter school appears smaller, it has more students per teacher. To compound matters, many studies consider a student-teacher ratio as a unit of measurement to distinguish schools with small-class sizes from those with larger class sizes but are often confusing class size with student-teacher ratio. According to Tienken and Achilles (2009) class size can only, truly, be determined by counting the students who actually sit in the teacher's class, not simply by dividing the total number of certified educators working in the school by number of students. In many cases, several certified staff members may not actually work in the classrooms and have limited exposure to the students. Class size is an organizational and logistical arrangement designed for instruction and teacher-student ratio is mainly a formula for resource allocation. The two are not the same and should be stringently considered thus, when evaluating research results (Tienken & Achilles, 2009).

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Many researchers have correctly considered the difference between class size and schools size and have rigorously studied the effect of class size on student achievement. Konstantopoulos (2008) evaluated data from 11,000 elementary school students in grades kindergarten to grade 3 and found that even the higher-achieving students benefited from small classes in early grades. The author indicated that all types of students benefit from small classes. Leahy (2006) found compelling evidence in a mixed-methods' study using a survey to gauge teacher opinions of class size, and the data revealed that most teachers either strongly agreed or agreed that smaller class sizes increase student achievement in many ways. For example, all Leahy (2006) respondents agreed large class sizes decrease student achievement, large class sizes have more discipline problems, and smaller classes sizes lead to faster gains in reading and math.

Other class-size studies have evaluated potential correlations between class size and student achievement. Bowne, Magnuson, Schindler, Duncan and Yoshikawa (2017) evaluated over 60 years of data from a comprehensive U.S. early childhood education program for their study. These authors claimed smaller teacher-student ratios were positively related to cognitive and achievement outcomes in classrooms, which corresponds with other studies that have suggested that these ratios may play a role in facilitating high-quality interactions between teachers, both those who are emotionally supportive and cognitively stimulating (Bowne et al.). Viadero (2008) reported on a British study of elementary and secondary students and their ability to stay on task, and found that students tended to be off task the larger the class size. Adding five students to a class decreased the odds of students being off task by nearly 25%. Additionally, low-performing students were almost twice as likely to be disengaged in a classroom with 30

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

students as they were in a classroom of 15 (Viadero, 2008). Teachers in larger classes are spending more time managing student behavior rather than engaging students.

Using global comparisons, U.S. teachers average 27 students per class versus the world average of 24 (Darling-Hammond, 2015). She suggested these large class sizes created compounding affects where teachers in the U.S. spend more time on direct instruction (27 hours per week versus 19), work more total hours each week (45 hours versus 38) and have less time in their schedules for planning, collaboration, and professional development. As a consequence, teachers lack adequate time to work with colleagues on creating compelling lessons, collaborating on new teaching methods, correcting student work, working with students individually, or consulting with parents (Darling-Hammond, 2015).

Simply reducing class sizes does not necessarily indicate increased student achievement or the attainment of other positive correlations. In the most ambitious of studies regarding class-size reduction, Florida and California invested significant legislative action and massive amounts of money to reduce class sizes. For instance, Florida invested billions of dollars to create a 1-to-20 student-teacher ratio for grades one through three (Januszka & Dixon-Krauss, 2008, p. 168). Across the country, in 1996, California enacted a similar, and just as costly, class-size reduction initiative. The program reduced class sizes, on average, from 30 to 20 students, in grades kindergarten through third grade (Jepsen & Rivkin, 2013). In both studies, modest gains were noticed in regards to student achievement, but many researchers questioned the cost benefit analysis of those gains (Januszka & Dixon-Krauss, 2008). In California for instance, many questioned the validity of decreasing class sizes at the expense of hiring



## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

inexperienced and unqualified teachers (Jepsen & Rivkin, 2013). Regardless of the numerous variables associated with class-size reduction initiatives, many researchers cannot deny the potential benefits of smaller class sizes.

### **PA System of School Assessment (PSSA)**

Starting in 2003, the PA Department of Education (PDE) began laying the groundwork for their new standardized testing method known as the PA PSSA. The process started with the development of Mathematics and Reading assessments for grades 5, 8, and 11. These items were field-tested in 2004 and launched in the spring of 2005. In 2004, the PDE began test development for additional grade levels in grades 4, 6, and 7, field-tested the assessments a year later, and fully implemented the testing in the spring of 2006. In 2007, a Grade 3 Mathematics and Reading assessment was developed through a similar process. In 2007, the Data Recognition Corporation (DRC) became responsible for the design, implementation, and scoring of the PSSA exams and has held the responsibility since. By 2008, students in Grades 5, 8, and 11 were taking Writing exams, and students in Grades 4, 8, and 11 were taking Science exams (PDE, 2017).

With the exception of shifts in test-taking windows, the PSSA exams remained relatively unchanged from 2009 through 2012. However, a transition to the PA Core Standards (PCS) in mathematics and English Language Arts (ELA) started taking form, and a shift to Keystone Exams, an end-of-the-year test, eliminated 11<sup>th</sup>-grade students from taking PSSA exams. Beginning with field-test questions in primary grades in 2013 and shifting to fully functional field sections for all grades in 2014, the completed transition to the PCS was accomplished by 2015 (PDE, 2017).

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

The PCS were based on Common Core Standards, and committees of educators were challenged to review, write, and approve the PSSA assessment anchors based on the new standards. After presentation of content materials and a basic blueprint structure, members drafted eligible content language and created vertically aligned curriculum across grade levels. The results of the committee were reviewed by local, state, and national subject experts, and, eventually, were reviewed by another committee. The standards were ultimately approved by the state board of education in the fall of 2013 (PDE, 2017).

As part of the PCS transition, legacy PSSA materials and content were completely phased out. Specifically, writing tests in all grade levels were eliminated and replaced with ELA exams. PSSA exams were no longer administered by alternating subjects, but now had a distinct math and reading section. ELA exams also included a text dependent analysis (TDA) section. TDA questions require substantial student writing and the ability to synthesize information from a passage. Where legacy writing responses were simpler, required less synthesis, and needed less explanation, the TDA components required students to develop a comprehensive, essay response. The committees believed the high demands of TDA questions coincided with demands required for students to be college and career ready, a major premise behind PCS (PDE, 2017).

The shift to PCS in 2015 caused a major statistical disruption compared to legacy exams. For math, the shift to PCS eligible content created a significant drop in state average scores. In 2014, the last year of legacy content, nearly 73% of PA students scored proficient or advanced. In 2015, the first year of the transition, only 40% of students were proficient or advanced. Since the reading and writing exams were

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

combined into one ELA exam, comparison between the results was difficult, but nevertheless, 70% of students were advanced and proficient in 2014 compared to 60% in 2015. Overall scores were so majorly affected by the shift to core standards, that, although scores were made public, the PDE decided not to allow the result to affect district profile scores or report cards (PDE, 2017).

In early December 2017, PA Governor, Tom Wolf, announced additional shifts in PSSA testing. Starting in the 2018 testing year, PSSA exams were condensed to eliminate at least two days of testing for most schools. Specifically, ELA exams were shortened from four sections to three, and math exams were reduced from three sections to two. This move has allowed the PDE to condense and move the testing window to later in the year starting with the 2019 school year and beyond (Governor Press Release, 2017). The decrease in the amount of testing may have eased the burden on students and teachers, but it did little to improve scores. Considering math and ELA averages, per grade, across the state, the scores did not shift more than three percentage points, and, in fact, many of the 2017 scores were higher than the shortened 2018 testing window (PDE, 2017).

A broad range of students participate in both the math and ELA PSSA exams including students from all public schools (nearly 800,000 students per test): a small amount (less than 3,000 students per test) of non-public school students who electively participate, home-schooled students (fewer than 600 students per test), and a small number of foreign exchange students (generally less than 200 students per test). Students who have significant cognitive impairments may be eligible to take an alternative exam known as the PA Alternate System of Assessment (PASA). Eligible students:

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

must be enrolled in the assessed grade level for the subject area, have a severe cognitive disability, require intensive instruction, require extensive adaptation and support to perform or participate meaningfully, require substantial modification of the general education curriculum, and participate in the general education curriculum that differs markedly, in form and substance, from that of other students. PASA students account for approximately 1% of all test takers (PDE, 2017, n.p.).

Over 20,000 students per year do not take PSSA exams for a variety of reasons, the largest percentage of which are comprised of parents who opt out students for religious reasons (42.5%), or for other reasons (14.55%). Other reasons include non-attempt (19%), ELL students during their first year of school (10%), medical emergencies (5.4%), other reasons (5%), extended absences from school (3.1%), and absent without completing makeup exam (2%) (PDE, 2017).

### **PSSA Reliability and Validity**

On some school-rating systems, standardized tests account for 80% of the public score; hence, the exams have high stakes for districts (PDE, 2017). Many curricular and academic decisions are based on the results, and questions regarding a student's ability to meet academic expectations are highly reliant on test performance. Evaluating the effectiveness of school intervention, and in this case school size, trusting the results of the exams is crucial. Also understanding the steps taken by exam designers to minimize testing bias for various subgroups is integral for district leaders and policymakers. Although the PDE develops processes to ensure the reliability of their testing procedures, many researchers have made contentions about the disadvantages of standardized test

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

design. Further exploring these questions is critical to understanding the concept of school size and student/subgroup achievement.

PSSA exams utilize a universally designed assessment developed by the National Center for Educational Outcomes (NCEO). The principles of such design allow for participation from a wide range of students and are based on the premise that “testing results should not be affected by disability, gender, race, or English language ability ([DRC], 1978, n.p.).” Data Recognition Corporation ([DRC], 1978) utilizes these design principles at every stage of item and test development to ensure that test materials are congruent with universal design. Their test development team was divided into varying content committees and were trained directly by NCEO. Some committees were involved in an intensive content review regarding the unique needs of students with disabilities and ELLs, and others were trained in bias, fairness, and sensitivity (PDE, 2017).

The most crucial elements of universally designed assessments include the following guiding principles developed by the NCEO and can be found in Appendix A. Item writers and reviewers used a checklist during the question development process to ensure every aspect of universal design was closely followed (PDE, 2017). The guidelines served as the central theme of the checklist and can be found in Appendix B.

The DRC (1978) takes several steps to maximize accessibility for all students across all content levels, including utilizing easily readable fonts and font sizes, as well as limiting shading, graphics, charts, and the number of items per page. When applicable, test-makers appropriately position graphics, pictures, diagrams, charts, and tables to be near their associated test questions. In addition to using high-contrast text to highlight

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

pertinent information, tests were also published with a dull finish to avoid the glare typically associated with glossy paper. Tests were also printed and bound in a manner to allow test booklets to lay flat for two-page viewing and to improve reading and handling of test materials (PDE, 2017).

While strictly adhering to universally designed assessments to ensure the participation of a vast range of test-takers, many students still required accommodations. The DRC (1978) allowed for accommodations to:

ensure students are not subjected to disadvantages during testing and that accommodations potentially utilized during regular instruction are still available to those students. Since literature related to assessment accommodations is still evolving and rarely provides empirical data that support the reliability and validity of the accommodations; PDE annually evaluates accommodation policies and current research assertions ([DRC], 1978, n.p.).

Although extensive steps are taken to ensure universal design and accessibility for all students, the DRC (1978) goes to great lengths to evaluate the reliability of the PSSA exams. Reliability is the property of a set of scores, not necessarily the assessment that produced the scores (Frisbie, 2005). In the 2014 version of the Standards for Educational Testing, reliability was considered the consistency of scores across various replications of testing procedures. According the PDE Technical Report (2017), reliability related to tests scores and not particularly the test, but aspects of scores could be affected by characteristics of the test. For instance, tests with more questions tend to be more reliable than tests with fewer questions. Furthermore, reliability tends to be higher in heterogeneous populations than more homogeneous populations.

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

The 2017 PSSA Technical Report noted a few concerns with reliability, but generally accepts PSSA exams as being a reliable measurement of achievement. While freedom from measurement error is quite important, reliability is associated with:

random sources of error and the degree of inconsistency due to random error.

However, systematic error may also exist which can artificially increase reliability. Also, the DRC (1978) noted that multiple sources of error could exist, including the day of testing, question usage, open-ended question raters, as well as others. Although multiple frailties exist in regard to reliability, many reliability indices simply reflect a single type of error; in turn, test users should understand what type of error is specifically being considered (PSSA Technical Report, 2017, n.p.).

The DRC (1978) also marked the distinction between relative error and absolute error and further indicated that reliability indices usually reflect only relative error (PSSA Technical Report, 2017).

As Frisbie (2005) suggested, reliability is a vast and complex notion and it cannot be adequately represented by a single number. Although several reliability measures are available, these may not provide the same results. The DRC (1978) utilized several evaluative techniques to ensure a reliable evaluation of student achievement including “reliability coefficients and their interpretation, unconditional and conditional standard errors of measurements, decisions’ consistency, and rater agreement” (PDE, 2017, n.p.).

As defined in the 2014 version of The Standards for Educational Psychological Testing, validity refers to the degree to which evidence supports the interpretation of test scores supported by the proposed utilization of the test. The sources referenced in the

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Standards include test content, response processes, the internal structures of the test, the relationships between tests scores and other variables, and the consequence of testing (The Standards for Education Psychological Testing, 2014, n.p.). Furthermore, the DRC (1978) utilized Item Response Theory (IRT) models to analyze the assessment data.

According to the PSSA Technical Report (2017), test content validity was highly reliant on establishing a link between test items and the assessment anchors and eligible content. In addition to this alignment, Lane (2014) suggested that content validity for standardized assessments should be measured by the degree that test items are representative, that panels of experts evaluate content, fairness reviews are conducted, testing procedures are evaluated, and third-party, independent reviews occur. Chapters two through eight of the 2017 PSSA Technical report extensively discuss the steps taken by the DRC (1978) to ensure these content validity standards are met.

The DRC (1978) utilized a Depth of Knowledge (DOK) model to ensure the PSSA items were aligned with Eligible Content and Academic Content standards both in regard to cognitive level and content. Additionally, the DRC (1978) created particular development specifications for items and determined the items were sufficient in number and were fairly distributed across content, cognitive ability, and difficulty. The DRC (1978) created minimum qualifications for item writers and provided extensive training to ensure they wrote high-quality questions. For any new-developed item, content specialists and editors made sure all items accurately measured the intended assessment anchors. Content specialists and editors also considered depth of knowledge, graphics, grammar, punctuation, language demand, and distractor reasonableness.



## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Even prior to field-test questions, items were submitted to content committees comprised of PA teachers to review the following information:

- Overall quality and clarity;
- Anchor, eligible content, and/or standard alignment;
- Grade-level appropriateness;
- Difficulty level;
- Depth of knowledge;
- Appropriate sources of challenge (e.g., unintended content and skills);
- Correct answer;
- Quality of distractors;
- Graphics; and
- Appropriate language (PDE, 2017, n.p.)

The DRC (1978) also created a Bias, Fairness, and Sensitivity Committee to review items as they related to diversity and gender. If items passed these standards, they were placed in a field test and several statistical analyses were conducted to determine the viability of the item. Items were once again reviewed by DRC (1978) officials and PA teachers, and differential item functioning (DIF) was used to detect items that might bias test scores for any specific group. Investigating the empirical results of DIF strengthens the evidence for validity related to score interpretations for particular groups. Finally, raters who evaluated open-ended questions were carefully recruited, screened, and trained and their scoring was highly monitored to ensure an acceptable level of scoring accuracy was achieved (PDE, 2017).

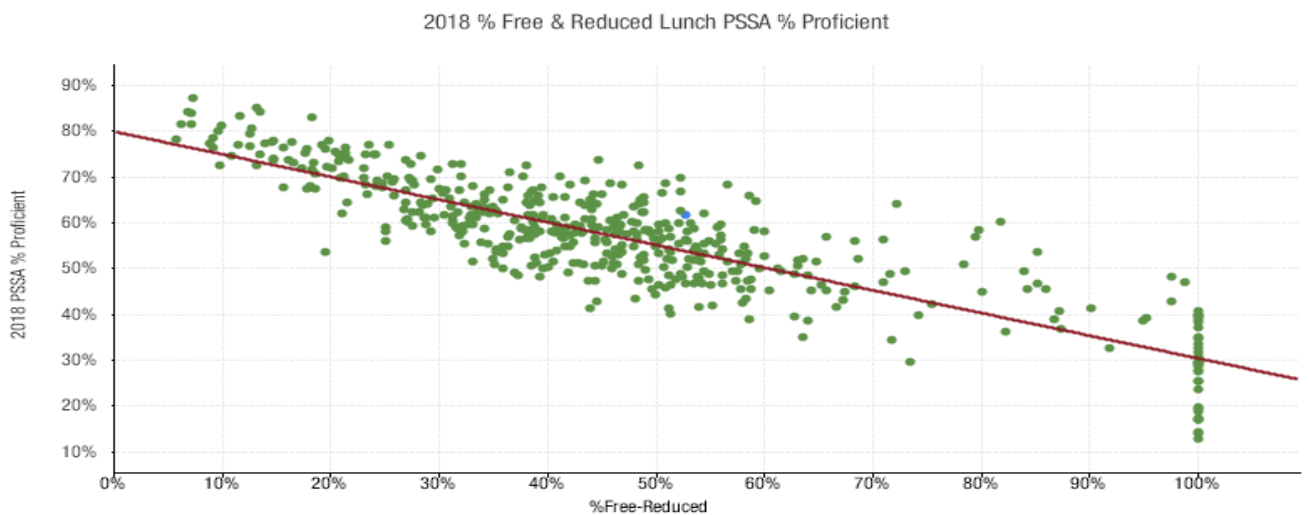
## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

In further investigation of content validity, the PDE also conducted a third-party review of assessment anchors, eligible content, and test items. In 2005, Achieve Inc. conducted a detailed investigation of content alignment and produced a report regarding their findings. Achieve Inc. (2005) is a bipartisan, non-profit organization designed to help states evaluate their academic standards, improve standardized assessments, and strengthen accountability, and, according to Achieve Inc.'s 2005 executive report, their mission is to provide state policymakers with an independent expert review of the quality of their standards and tests (Achieve Inc., 2005). Achieve Inc. (2005) concluded that PA had generally identified the most essential content for inclusion in the assessment anchors and eligible content statements. Researchers went on to say that assessment anchors in reading and mathematics align well with PA's academic standards, and eligible content statements are clear and measurable. The general steps referred to, above, and the extensive research reports on curriculum alignment are both examples of PA's insistence on the content validity of the state achievement tests and their alignment to state standards.

In another check for validity, the DRC (1978) endeavored to examine the extent to which the cognitive skills of students match the exams' defined-construct domains. Known as response-process evidence, test developers utilized think-aloud procedures or cognitive labs to assess the congruency. Furthermore, open-ended questions were evaluated regarding the raters' ability to interpret and apply scoring criteria to student answers (PDE, 2017). According to the DRC (1978), all scorers were subjected to a "well-organized training and strict monitoring of rating accuracy," all of which were designed to minimize scorer bias (n.p.).

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Although several mechanisms are in place to minimize bias and have a reliable and valid testing instrument, the literature reviewed earlier does indicate strong correlations to testing performance and several demographic variables, including SES, race, and special education status. As that data have been well established around the country and the world, PSSA data indicate the same type of patterns. Figure 1 shows the correlation of free- and reduced-lunch rate versus the percent proficient on the PSSA exams by school district. As the free- and reduced-lunch rate increases, the percent proficient on the exams decreases as well.



*Figure 1.* Percent of students qualifying for free and reduced lunch versus the percentage proficient on PSSA exams by district (PDE).

### **PA Value-Added Assessment System (PVAAS)**

Since the inception of standardized testing in PA, educators and policymakers have been provided a tool that spans across subjects and grades to provide two crucial pieces of information. First, PVAAS provides information on the academic growth students have made in the most recently assessment school year. Growth, in this sense, refers to the academic progress groups of students have made, and educators can utilize

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

the information to assess the success of their academic preparation and instructional strategies. Second, PVAAS provides information about the student's potential individual success on standardized tests. In other words, PVAAS provides a prediction about a student's success on the PSSA . According to the Statistical Analysis System Educational Value-Added Assessment System ([SAS EVAAS], 2016), the value-added data is based on a reliable and robust methodology, designed to overcome statistical issues related to standardized testing and minimize concerns regarding fairness.

To help understand the distinct difference between achievement and growth, PDE (2018) utilized an analogy where a two track coaches are trying to evaluate the progress of their respective teams. Team A has never finished in the top half of all teams at their track meets; however, Team B has always placed near the top three spots. Although Team B has experienced more success or achievement than Team A, both coaches likely want their teams to grow and improve. Regardless of their team success, both coaches have an ability to evaluate the growth of their team. In this case, each coach would utilize the times from all races, over a period of time, to determine where each team started and how the team progressed. Reasonable goals for either coach could be to have their team maintain their speed over time or improve their team, and, obviously, neither coach would want their team to lose ground. Perhaps Team A, although highly successful, can improve their team and even make enough progress to break their team record, and, although Team B may not be able to win a meet, they can show improvement from their times earlier in the season or previous seasons. Overall, regardless of their success, each coach can determine if his team is improving, maintaining, or regressing.

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

This analogy of track teams is connected to academic growth measurement for groups of students. Students who are similar to Team B demonstrate high achievement, scoring proficient or advanced on the PSSA exams, but can still show improvement over team, or, can actually not show improvement. Students similar to Team A may not be high achieving, but they can quite possibly demonstrate growth and improvement. They can also not show improvement or growth. The concept that students can show improvement regardless of their achievement is exactly why researchers often claim that growth is not affected by student demographics or membership in any particular subgroup (PDE, 2018).

To help further dissect PVAAS, data were collected from the state assessments regarding groups of students from year to year. In other words, PVAAS does not compare the results of one group of students in a specific subject and grade with the previous year's students in the same subject and grade. For instance, the achievement for the most recent group of students in 5<sup>th</sup>-grade math is not compared to the achievement of the previous year's group of 5<sup>th</sup>-grade math. The most recent 5<sup>th</sup>-grade students will be compared to their achievement results from 3<sup>rd</sup>- and 4<sup>th</sup>-grade. PVAAS does not utilize the percentages of students at various performance levels, but it evaluates the range of scores in each performance group, especially considering how students fluctuate between ranges (PDE, 2018).

The growth is determined in an assessment against the standard for PA Academic Growth, which according to PVAAS, was based on the philosophy that regardless of the entering level of achievement for a group of students, no group should lose ground, academically. Growth ratings are displayed in the form of a color system. Red represents

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

strong evidence that students did not meet academic growth standards, and yellow indicates that moderate evidence exists that students did not meet academic growth. Green shows that students met or maintained academic growth, where light blue indicates that moderate evidence exists to show that students exceeded academic growth, and dark blue shows that strong evidence exists that students exceeded established growth standards (PDE, 2018).

Understanding the difference between growth and achievement data is crucial, but it is important to determine if PVAAS calculations are a reliable and valid form of measurement, and, to assess any potential weaknesses in its design. Although according to SAS EVAAS (2016), PVAAS utilizes a sophisticated and robust modeling approach to minimize concerns regarding the data, but many educators are concerned that PVAAS is reliant on the standardized tests scores, which have inherent limitations, themselves. SAS EVAAS (2016) does recognize that student test scores are the basic ingredient of PVAAS' analyses, but researchers asserted that PA assessments are statistically vetted and PVAAS processes are sound.

SAS EVAAS (2016) contended that, although standardized tests could create concerns surrounding fairness, PSSA exams were created by universal design, and their output data were:

sufficient for longitudinal modeling and prediction. Regardless, before utilizing any test in PVAAS calculations, rigorous analyses existed to verify the tests met the following three criteria: exams are sufficiently aligned to curriculum standards, must be reliable and valid, and demonstrate sufficient stretch at the extremes (n.p.).

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

According to SAS EVASS (2016), PSSA exams have met these criteria, and researchers have confirmed there was enough difference in:

scores at the top and bottom of the scales to differentiate growth from student achievement. After careful analysis of the percentages of students scoring at the top and bottom scores, researchers determined there are no ceilings or floors, and both high- and low-achieving students can realize both high- and low-growth (n.p.).

While concerns may arise about a narrow snapshot of scores, PVAAS utilizes many tests scores across subjects, grades, and years (SAS EVAAS, 2016).

Since much of the literature discussed thus far indicates that students in particular demographics or socioeconomic statuses often score lower on exams or experience biases during testing, determining if PVAAS is an accurate method of verifying growth is a crucial component. Although SAS EVAAS (2016) admitted that certain subgroups scored lower on standardized exams, researchers also asserted the disadvantages were minimal as it related to PVAAS' reporting. In fact, researchers explained the sophisticated value-added models utilized in PA mitigated this concern by using "all available testing history for each individual student, and in essence, each student served as his or her own control." (SAS EVAAS, 2016, n.p.) The influence of SES or other demographics tend to persist over time; therefore, these influences are already represented in the students' data. Lockwood and McCaffrey (2007) presented analytical and simulation data that suggested that value-added processes are extremely effective at purging student heterogeneity and compress biases of traditional standardized tests.

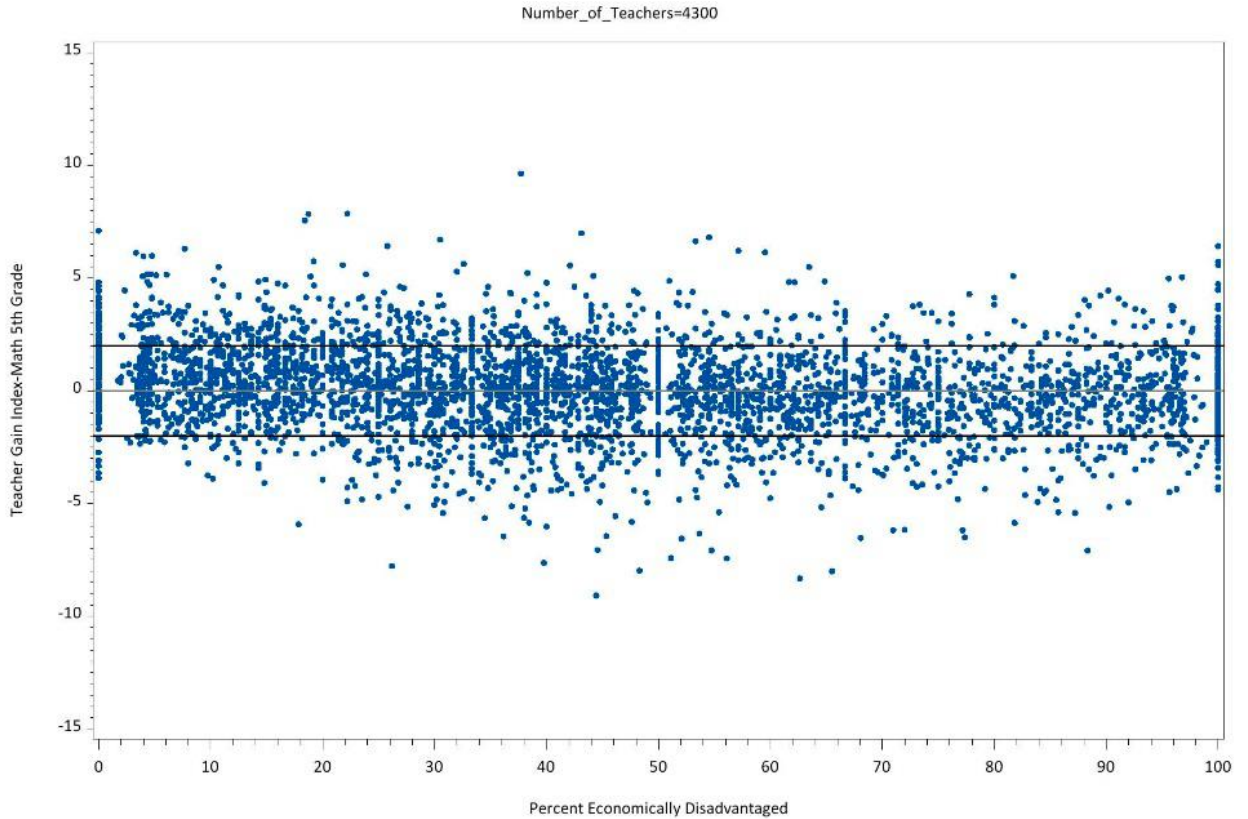
## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

In a study designed to explore models of student performance, Choi, Goldschmidt, and Yamashiro (2006) indicated their correlations demonstrated important findings regarding value-added theory and adjustments for student and demographic characteristics. When controlling for SES, the value-added model captured the preceding effects that SES might have on students. Because of these correlations, results of value-added models lead to more valid inferences because the models are accounting for differences in initial academic status and moderate factors such as SES or language proficiency. Additionally, researchers noted that value-added models provided the most valid and informative picture of student and school performance (Choi et al.).

SAS EVAAS (2016) also believed PA's own value-added data indicated neither an advantage or disadvantage existed for educators considering the vast type of students in classrooms. The graph in Figure 2 plots the percentage of students qualified as free/reduced lunch against teacher growth indices for PSSA Mathematics in 5<sup>th</sup> grade for testing year 2014. Regardless of the SES status, "there is essentially no correlation to the growth index (SAS EVAAS, 2016, n.p.)." More simplistically, "the dots representing each teacher do not trend up or down as the percentage of increase, and the cluster of dots remains relatively flat across the spectrum (SAS EVAAS, 2016, n.p.)." An actual .139 correlation between the growth index and SES indicates a weak relationship (SAS EVAAS, 2016).



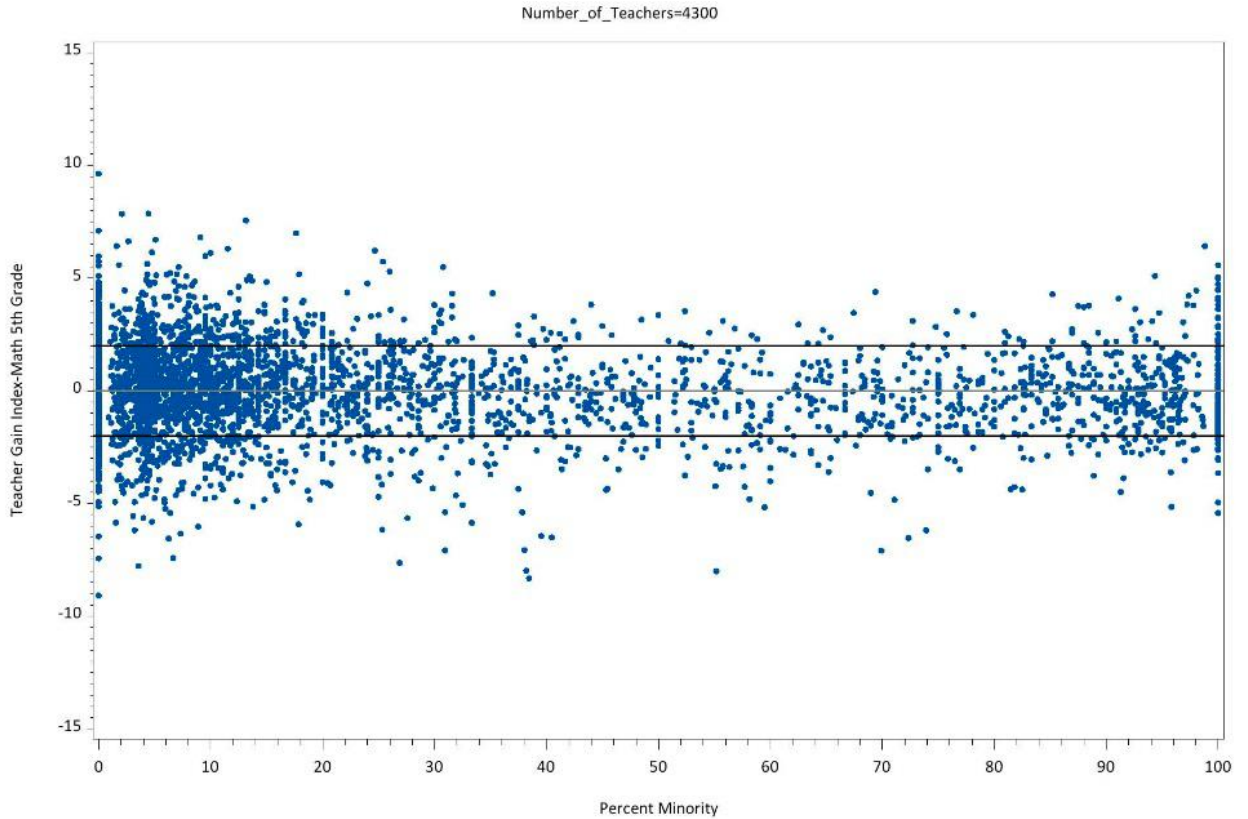
RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION



*Figure 2.* Percentage of students qualified as free/reduced lunch against teacher growth indexes for PSSA Mathematics in 5<sup>th</sup> grade for testing year 2014. Statistical Analysis System Educational Value-Added Assessment System. (2016). Retrieved from [https://www.sas.com/en\\_us/software/evaas.html](https://www.sas.com/en_us/software/evaas.html)

Figure 3 shows similar correlation for the percentage of minority students. The actual correlation is a negligible  $-0.070$ , which demonstrates essentially no correlation to the growth index and minority status.

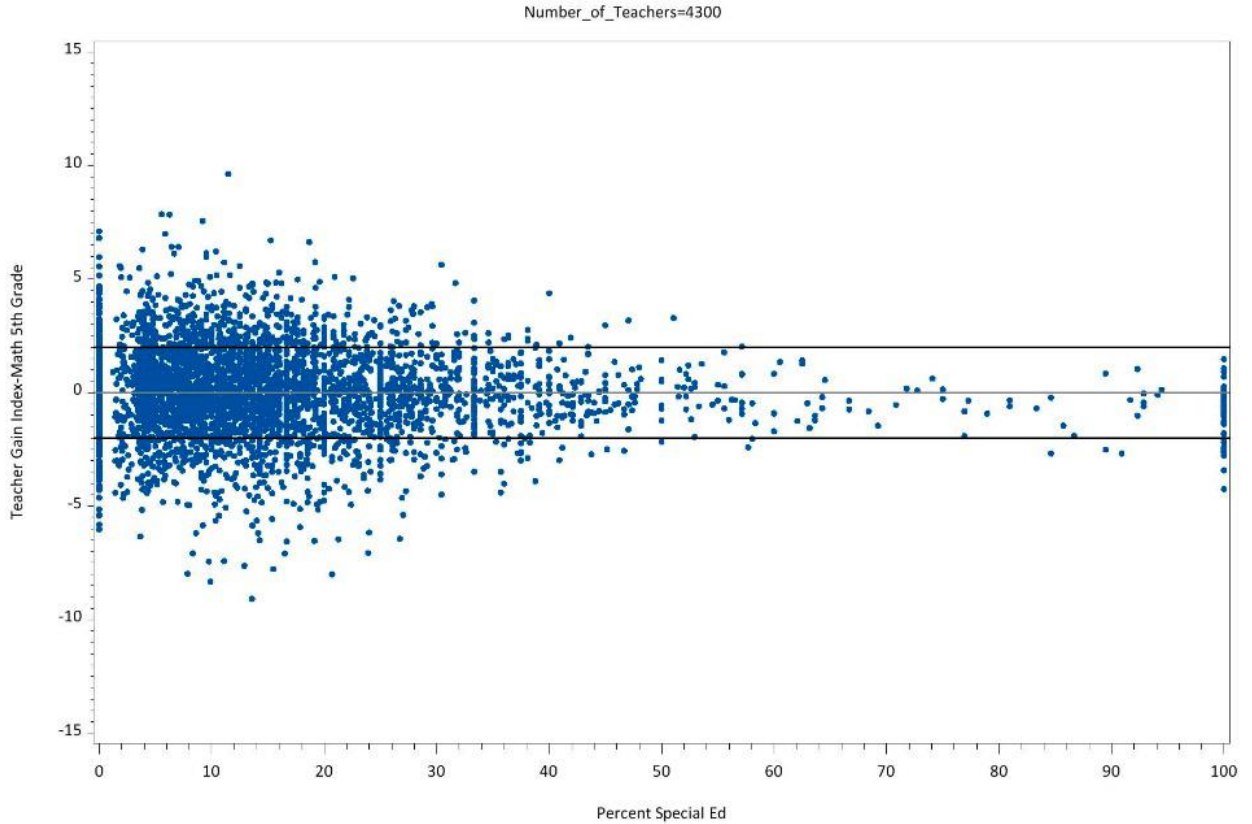
RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION



*Figure 3.* Percentage of students labeled as a minority status against teacher growth indexes for PSSA Mathematics in 5<sup>th</sup> grade for testing year 2014. Statistical Analysis System Educational Value-Added Assessment System. (2016). Retrieved from [https://www.sas.com/en\\_us/software/evaas.html](https://www.sas.com/en_us/software/evaas.html)

Figure 4 provides similar information, but it uses the percentage of students considered to have special education needs. The correlation between the growth index and percentage of special education students is .096; again, another indication of a weak relationship.

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION



*Figure 4.* Percentage of students qualifying for special education against teacher growth indexes for PSSA Mathematics in 5<sup>th</sup> grade for testing year 2014. Statistical Analysis System Educational Value-Added Assessment System. (2016). Retrieved from [https://www.sas.com/en\\_us/software/evaas.html](https://www.sas.com/en_us/software/evaas.html)

PSSA and PVAAS data have contrasting differences in data development and offer distinctly different information for teachers, schools, and districts. In the context of this study, evaluating the effect of school size on student achievement on the PSSA exams, specifically the results of subgroup statuses, should produce more valuable and pragmatic data. Although no one assessment system will alleviate statistical concerns or perfectly portray student information, value-added data in PA has been adequately vetted from a statistical standpoint. Furthermore, the weak correlations involving subgroups and growth data would likely hide any effect school size may have on subgroup performance.

### **Summary**

School size research has an extensive history, and the debate regarding the effectiveness of school size has raged on for decades. Various traditional, large-scale studies exist that contend smaller schools are more effective than larger schools, and more recent studies have tried to narrow in on the effects for contingents of students. Some researchers asserted that historically underperforming students, such as those who are economically disadvantaged, have a disability, or belong to a minority, demonstrate higher achievement in smaller schools compared to the same population in larger schools. On the other hand, some researchers have found mixed results or find the size of the school has little or no effect, regardless of the subgroup. The current study will attempt to determine the potential relationship of school size on student achievement for historically underperforming subgroups utilizing Pennsylvania standardized tests in grades three through eight.

Chapter 3

**Methodology**

The purpose of this quantitative study is to determine whether a relationship exists between school size and student achievement; specifically, whether a relationship exists based on student performance on the PSSA exams in grades 3-8 for math and reading. A request was made to the PA DOE and the Bureau of Curriculum, Assessment, and Instruction for PSSA results in the subjects of math (grades 3-8) and ELA (grades 3-8). The request was made to receive the data in a particular format, but all information provided was publicly available. The achievement of students who are economically disadvantaged, receive special education services, and/or belong to a minority racial category helped or hindered by school size is also examined. The study also evaluates whether these variables along with school size were able to predict a significant amount of variance in student achievement.

**Research Questions**

The research questions for this study are as follows:

1. Do students who receive free/reduced lunch demonstrate significantly higher student achievement on standardized tests in smaller schools compared to larger schools?
2. Do students who have an IEP demonstrate significantly higher student achievement on standardized tests in smaller schools compared to larger schools?
3. Do students who belong to a minority racial status demonstrate significantly higher student achievement on standardized tests in smaller schools compared to larger schools?

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

### Participant Data

Student achievement results, economically disadvantaged status, IEP rates, minority enrollment and school size was compiled via the information request to the PDE Bureau of Curriculum, Assessment, and Instruction. Data was collected on all 500 public school districts in Pennsylvania using 2015 through 2018 data. The data request was conducted for the purpose of receiving the information in a particular format, although all of the information provided is publicly available on the PDE website. Percentage of proficiency was determined by adding the number of students scoring advanced and the number of student scoring proficient and dividing by the total number of students participating. A snapshot of the actual data set as it relates to school size and economically disadvantaged rates are displayed in Table 1.

Table 1.

*School Level PSSA Data by Subgroup and Enrollment*

<u>District</u>	<u>School</u>	<u>Subject</u>	<u>Subgroup</u>	<u>Grade</u>	<u>Enrollment</u>	<u>%Proficient</u>
Abington	Abington MS	ELA	ED	5	236	45.2
Albert	North MS	ELA	ED	5	134	67.8
Aliquippa	Aliquippa Elementary	ELA	ED	5	82	21.8

### Data Sources

Student achievement results, economically disadvantaged status, IEP rates, minority enrollment and school size was compiled via the information request to the PDE Bureau of Curriculum, Assessment, and Instruction. Data was collected on all 500 public school districts in Pennsylvania using 2015 through 2018 data. As stated previously, this window of data was utilized mainly because of the major shift in the data produced prior

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

to 2015, following the shift to Pennsylvania Common Core standards. A data request was conducted for the purpose of receiving the information in a specified format, although all of the information provided is publicly available on the PDE website.

### **Data Collection Procedures**

After receiving approval from the Youngstown State University (YSU) Institutional Review Board (IRB), the PDE Bureau of Curriculum, Assessment, and Instruction was contacted to request the student and school data. The data is comprised of student achievement results on the reading and math PSSA exams, economically disadvantaged status, IEP rates, minority enrollment and school size. Data were collected on all 500 public school districts in Pennsylvania using 2015 through 2018 data in grades 3 through 8. The data were sent via email in the form of several Microsoft Excel spread sheets, and were eventually consolidated into one spreadsheet for analysis. Schools were identified in the spread sheet for the purpose of grouping schools by size; however, all the information provided in the data base is publicly available.

### **Method of Analysis**

The proposed data analysis for the current investigation is a primary data analysis (PDA) meta-analysis. Because the data were examining the number of students testing and not testing proficient across the different groups, the primary PDA analytical approach uses odds ratios. The odds ratio approach produced a basic-effect size measure for the entire data set (between those testing/testing proficient) based on a determined reference group and provided specific effect-size estimates for each level of the data (school size, economically disadvantaged status, IEP rates, minority enrollment). An odds-ratio value equal to or approximating one indicated no differences between the

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

reference group and the comparison group on passage rates. An odds' ratio value exceeding one indicates the reference group is out-performing the comparison group; an odds' ratio value below one is underperforming in relation to the comparison group. In addition to the effect-size estimates, the confidence intervals and significance tests were produced for all levels of the data, examining both the within group effects (i.e., across all minorities) and the between group effects that present. Individual primary data level (IPD) meta-regression will be incorporated if the data support this analysis.

### **Limitations**

Although a need exists to evaluate PA schools, the external validity of this study would be suspect outside of the sample. Using the free- and reduced-lunch rates is a fairly vast metric, and it does not necessarily offer potentially valuable details. For example, families may qualify for free and reduced lunches with annual salaries 100% to 180% above the poverty line. Determining the differences between the poorest students and those who barely qualify may be useful, but impossible with this structure. The simplistic correlation and regression analysis does not allow for any school culture analysis or gather any qualitative data. As with any linear analysis, PSSA data would be sensitive to outliers.

Although the data may suggest a relationship exists between school size and student achievement, merely identifying the size of the school as the reason for higher or lower achievement would be premature and overly simplistic. As the literature review suggested, a myriad of factors have an influence on student achievement and many factors, including school size, may create the existence of those factors and the influence they have on student achievement. Issues including but not limited to school climate,



## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

teacher attitude and beliefs, inclusionary methods, and others may have a substantial impact on student achievement; however, the size of the school may increase/decrease the likelihood those and other factors exist. Considering this dynamic, the study may illuminate trends in regards to the relationship of school size and student achievement and may highlight the potential existence of external factors effecting achievement, but the scope of the study will not permit assertions on external factors nor will it allow for the elaboration on those factors.

### **Summary**

A significant amount of research exists regarding school size and student achievement, many of which utilize hierarchical linear modeling and correlation analysis. Discrepancies are apparent between studies in regard to school size and achievement for prototypical students, but some consensus exists with the positive effect small schools have on poorer, minority students. Although some researchers have seen little effect on student achievement in narrower studies, others have found evidence that students benefit from improved teacher attitudes and beliefs in smaller schools.

On many scholarly and political fronts, the debate rages on regarding school size and the effectiveness of school systems. Several studies indicate a relationship between school size and student achievement, many of which claim that poorer students are better served in smaller districts over larger. Although plagued with inherent limitations, this study offers valuable data for further educational planning and research in PA schools.

Chapter 4

**Results**

The purpose of this quantitative study is to determine whether a relationship exists between school size and student achievement; specifically, whether a relationship exists based on student performance on the PSSA exams in grades 3-8 for math and reading. A request was made to the PA DOE and the Bureau of Curriculum, Assessment, and Instruction for PSSA results in the subjects of grades 3-8 math and ELA. The request was made to receive the data in a particular format, but all information provided was publicly available. The achievement of students who are economically disadvantaged, receive special education services, and/or belong to a minority racial category helped or hindered by school size is also examined. The study also evaluates whether these variables along with school size were able to predict a significant amount of variance in student achievement.

**Research Questions**

The research questions for this study are as follows:

1. Do students who receive free/reduced lunch demonstrate significantly higher student achievement on standardized tests in smaller schools compared to larger schools?
2. Do students who have an IEP demonstrate significantly higher student achievement on standardized tests in smaller schools compared to larger schools?
3. Do students who belong to a minority racial status demonstrate significantly higher student achievement on standardized tests in smaller schools compared to larger schools?

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Prior to data analysis, 67,000 cases of school data that were provided through a public records' request were formatted to indicate the number of students passing and failing state level assessments in reading and mathematics for grades 3-8. The cases were aggregated according to their respective groupings (i.e., SES, IEP, Race, etc.) for the purposes of analysis. All data were imported into Comprehensive Meta-Analysis® so that all odds'-ratio values could be weighted according to the respective sample sizes available.

Fixed effect, odds'-ratio analyses were conducted to understand the potential influence of school size on student achievement. Results are presented initially for school size overall, grade-level overall, core area overall, and based on the variables of gender, SES, IEP, and race. Results for the overall effect-size measures indicate that based on  $n = 468$  aggregate school cases, a fixed-effect model suggests significant heterogeneity across all the aggregate cases,  $Q=826910$ , based on 467 df,  $p < .001$ . The results presented in Table 2 provide the odds'-ratio across the different school sizes.

RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Table 2.

*Overall Effect Estimate on School Size*

School Size	Point estimate	Lower limit	Upper limit	Z-value	P-value
≤ 100	2.40	2.39	2.41	553.24	<.001
101 - 200	2.27	2.26	2.28	422.18	<.001
201 - 300	2.27	2.26	2.29	325.92	<.001
301 - 400	2.22	2.21	2.24	226.55	<.001
401 - 500	2.09	2.06	2.12	118.01	<.001
≥ 501	2.24	2.21	2.27	109.28	<.001

As indicated above, historically underperforming students are at least twice more likely to fail their state level assessment relative to the non-historically underperforming students. Results indicate the within-effect size measures are significantly different overall,  $p < .001$ . Additionally, the between-effect size measures, indicating the size of the school, are significantly different,  $p < .001$ . However, while the between effects are significantly different, these differences are not practically different. More in depth analysis with school size is presented below. Table 3 provides the odds ratios across the different grade levels.

RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Table 3.

*Overall Effect Estimate by Grade*

Grade level	Point estimate	Lower limit	Upper limit	Z-value	P-value
3	2.48	2.47	2.50	372.25	<.001
4	2.45	2.44	2.47	363.77	<.001
5	2.46	2.45	2.48	362.25	<.001
6	2.32	2.31	2.33	334.45	<.001
7	2.18	2.16	2.19	303.31	<.001
8	1.98	1.97	1.99	258.35	<.001

Results indicate that the within-subjects effect size measures are significantly different overall,  $p < .001$ . Additionally, the between-subjects effect size measures are significantly different,  $p < .001$ . However, while the between effects are significantly different, these differences are not practically different. Table 4 will present effect-size measures by reading and math.

RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Table 4.

*Estimated Effect-Size Measures Across Reading and Math*

Subject	Point estimate	Lower limit	Upper limit	Z-value	P-value
Reading	2.26	2.25	2.27	577.52	<.001
Math	2.38	2.37	2.39	577.69	<.001

Results indicate that the within subject's, effect-size measures are significantly different overall,  $p < .001$ . Additionally, the between-subjects, effect-size measures are significantly different,  $p < .001$ . However, while the between effects are significantly different, these differences are not practically different. Table 5 will present effect-size measures by subgroup factors such as gender, IEP status, SES, and race.

Table 5.

*Estimated Effect-Size Measures by Subgroup Factor*

Factors	Point estimate	Lower limit	Upper limit	Z-value	P-value
Gender	0.88	0.87	0.88	-31.95	<.001
Individualized Education Plan	4.18	4.12	4.24	202.05	<.001
Socioeconomic Status	4.36	4.32	4.40	332.04	<.001
Race	4.19	4.15	4.23	289.09	<.001

RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Results indicate that the within-subjects, effect-size measures are significantly different overall,  $p < .001$ . Additionally, the between-subjects, effect-size measures are significantly different,  $p < .001$ . In particular, the subgroup factor of gender is where effect-size differences are present.

The next results present the odd's ratio for school size based on each subfactor based on gender.

Table 6.

*Estimated Effect-Size Measures on Gender by School Size*

Size Level	Point estimate	Lower limit	Upper limit	Z-value	P-value
≤ 100	0.75	0.74	0.77	-36.67	<.001
101 - 200	0.75	0.74	0.76	-35.75	<.001
201 - 300	0.78	0.76	0.79	-19.92	<.001
301 - 400	0.77	0.74	0.81	-11.42	<.001
401 - 500	0.61	0.55	0.67	-10.40	<.001
≥ 501	0.79	0.67	0.93	-2.86	<.001

The results indicate the reference group for gender “males” are passing exams 21% - 39% less often than female students. However, this analysis does not consider other subgroup factors, which will be addressed in future results.

Table 7 provides the estimated effect-size measures by school size and subject.

RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Table 7.

*Estimated Effect-Size Measures on School Size by Gender and Subject*

Size	Reading	Mathematics
$\leq 100$	0.61	0.97
101 - 200	0.58	0.96
201 - 300	0.59	0.99
301 - 400	0.60	0.98
401 - 500	0.58	0.91
$\geq 501$	0.67	0.97

The results indicate the reference group gender “males” are passing exams 33% to 42% less often than female students. However, male and female students pass the math exams at nearly equal rates. These results are presented graphically in Figure 5.

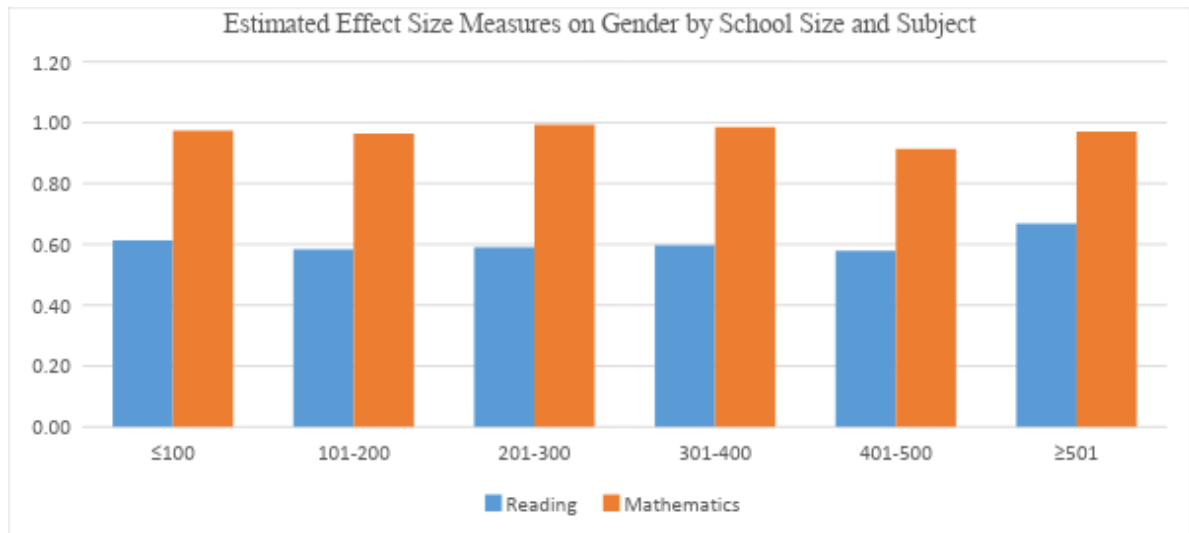


Figure 5. Shows the estimated effect size on gender by school size and subject.



RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Table 8 provides the results for IEP and school size.

Table 8.

*Estimated Effect-Size Measures on Individualized Education Plans by School Size*

Size Level	Point estimate	Lower limit	Upper limit	Z-value	P-value
≤ 100	6.22	5.98	6.47	89.40	<.001
101 - 200	7.58	7.37	7.81	136.09	<.001
201 - 300	7.82	7.50	8.16	96.46	<.001
301 - 400	6.94	6.47	7.45	53.96	<.001
401 - 500	9.27	8.03	10.70	30.47	<.001
≥ 501	8.99	5.49	14.71	8.74E+00	<.001

The results indicate the reference group of students without an IEP are more likely to pass state achievement exams. Results suggest this gap ranges from 6.2 to 9.3 times more likely to be successful for non-identified students. Additionally, the trend suggests that larger schools have larger gaps than smaller schools. Table 9 breaks out IEP results by subject area.

RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Table 9.

*Estimated Effect-Size Measures on Individualized Education Plans by School Size and Subject*

Size	Reading	Mathematics
≤ 100	6.46	5.64
101 - 200	8.24	6.58
201 - 300	8.65	6.86
301 - 400	7.74	6.10
401 - 500	9.28	9.10
≥ 501	8.75	10.42

Consistent with the results in Table 8, the larger the school the wider the gaps between students. Noteworthy is the gap between students with IEPs and students without, which is approximately half in smaller schools than in larger. These results are presented graphically in Figure 6.

RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

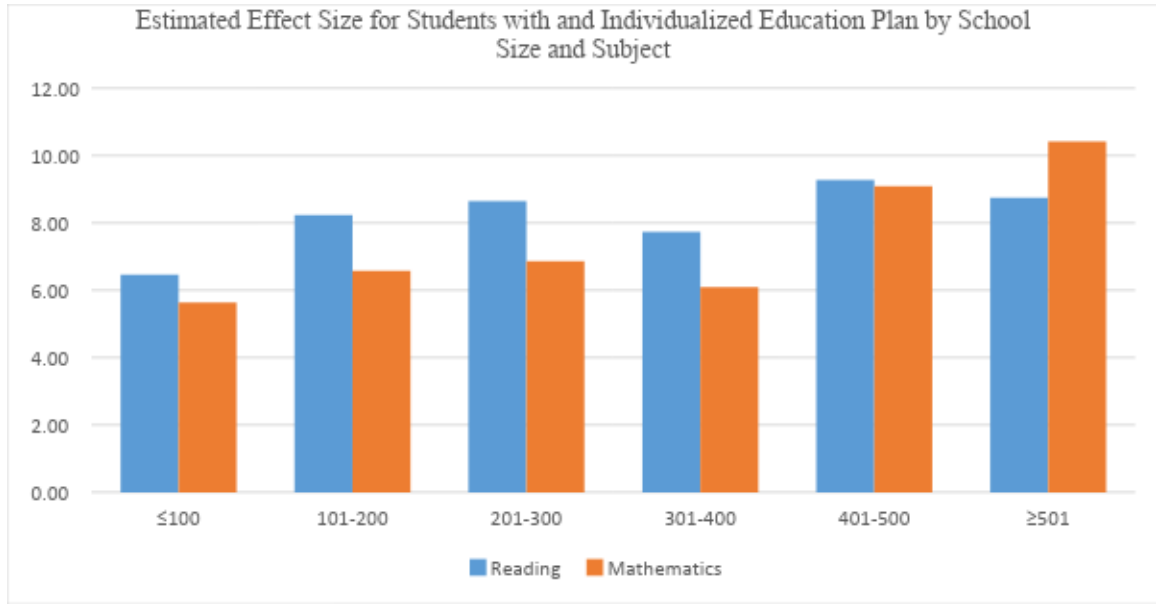


Figure 6. Shows the estimated effect size for students with an Individualized Education Plan by school size and subject.

Table 10 presents estimated effect-size measures for SES and school size.

Table 10.

*Estimated Effect-Size Measures on Socioeconomic Status by School Size*

Size Level	Point estimate	Lower limit	Upper limit	Z-value	P-value
≤ 100	4.00	3.94	4.06	170.45	<.001
101 - 200	3.93	3.87	4.00	161.64	<.001
201 - 300	4.11	4.00	4.23	103.06	<.001
301 - 400	4.31	4.10	4.52	59.08	<.001
401 - 500	4.49	4.16	4.85	38.56	<.001
≥ 501	4.90	4.57	5.27	43.64	<.001

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

The results indicated that students not receiving free and reduced lunches are 3.9 to 4.9 times more likely to pass state assessment exams than students receiving free and reduced lunches. Results also indicated an upward trend in the gap between the two groups as school size increases. These results are further broken out in Table 11 by subject area.

Table 11.

*Estimated Effect-Size Measures on Socioeconomic Status by School Size and Subject*

---

Size	Reading	Mathematics
≤ 100	3.97	4.04
101 - 200	3.89	3.97
201 - 300	4.12	4.10
301 - 400	4.26	4.35
401 - 500	4.54	4.45
≥ 501	4.64	5.23

---

Similarly with overall scores, the achievement gap increases between groups as school size gets larger. In particular, the gap in reading grows from 4.0 to 4.6, whereas, the gap for math increases from 4.0 to 5.2. The results are graphically displayed below in Figure 7.

RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

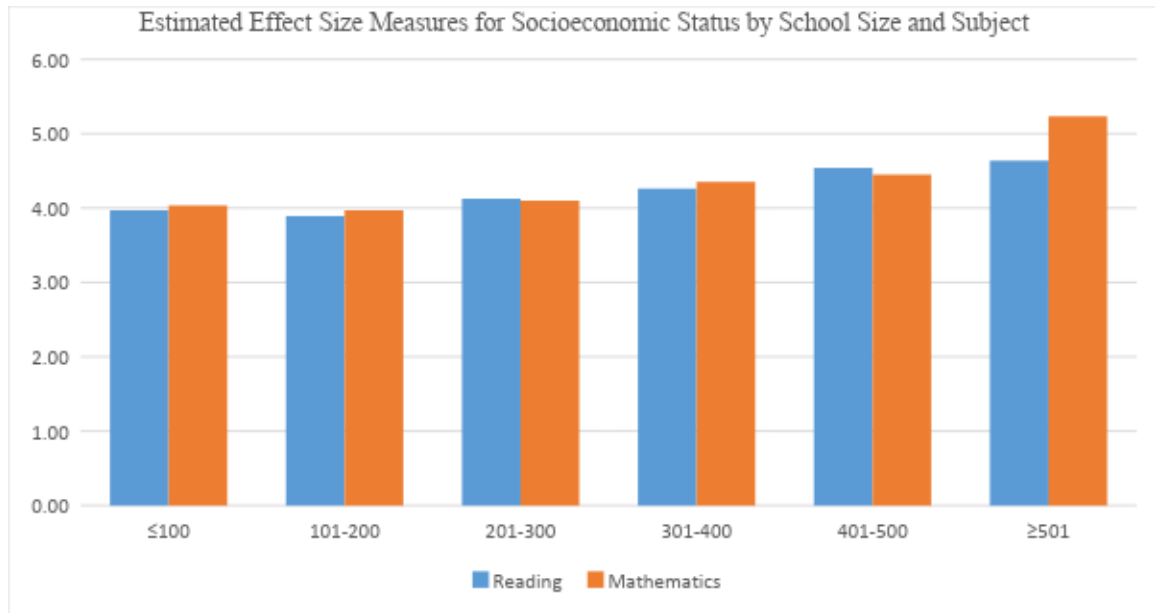


Figure 7. Shows estimated effect-size measures for Socioeconomic Status by school size and subject

Table 12 presents race overall by school size.

Table 12.

*Estimated Effect-Size Measure on Race Overall by School Size*

Size Level	Point estimate	Lower limit	Upper limit	Z-value	P-value
≤ 100	4.22	4.15	4.30	160.39	<.001
101 - 200	3.63	3.56	3.71	122.59	<.001
201 - 300	3.02	2.93	3.10	75.42	<.001
301 - 400	3.15	3.01	3.31	47.33	<.001
401 - 500	3.51	3.23	3.80	30.47	<.001
≥ 501	2.32	2.17	2.48	24.93	<.001

The results indicate the reference group of White students are more likely to pass state achievement exams, relative to the other groups (Black, Hispanic, Asian, Multi-

RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Race). The students identified as Native American and Pacific Islanders were not included in this analysis because of small sample sizes.

Table 13 provides effect-size measures by specific race and school size.

Table 13.

*Estimated Effect-Size Measures on Race Specificity by School Size*

---

Size	Asian	Black	Hispanic	Multi
≤ 100	0.61	5.31	4.45	2.19
101 - 200	0.53	5.19	4.30	2.35
201 - 300	0.33	4.82	3.73	2.03
301 - 400	0.41	4.78	3.20	1.80
401 - 500	0.47	6.81	3.26	1.53
≥ 501	0.29	4.01	3.78	1.97

---

As indicated, Asian students demonstrate substantially different results than other race categories. Asian students are more likely to pass state assessments than White students. However, Black, Hispanic, and Multi-Racial students are less likely to pass relative to White students. The achievement gap is largest for Black students, followed by Hispanic students, and finally, Multi-Racial students. The trend for Black students indicates the achievement gap narrows as the school size increases, with exception of the 401 - 500 size school. This trend is consistent with Hispanic and Multi-Racial students, with the

RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

exception of schools with 501 students or more. These results are graphically represented in Figure 8.

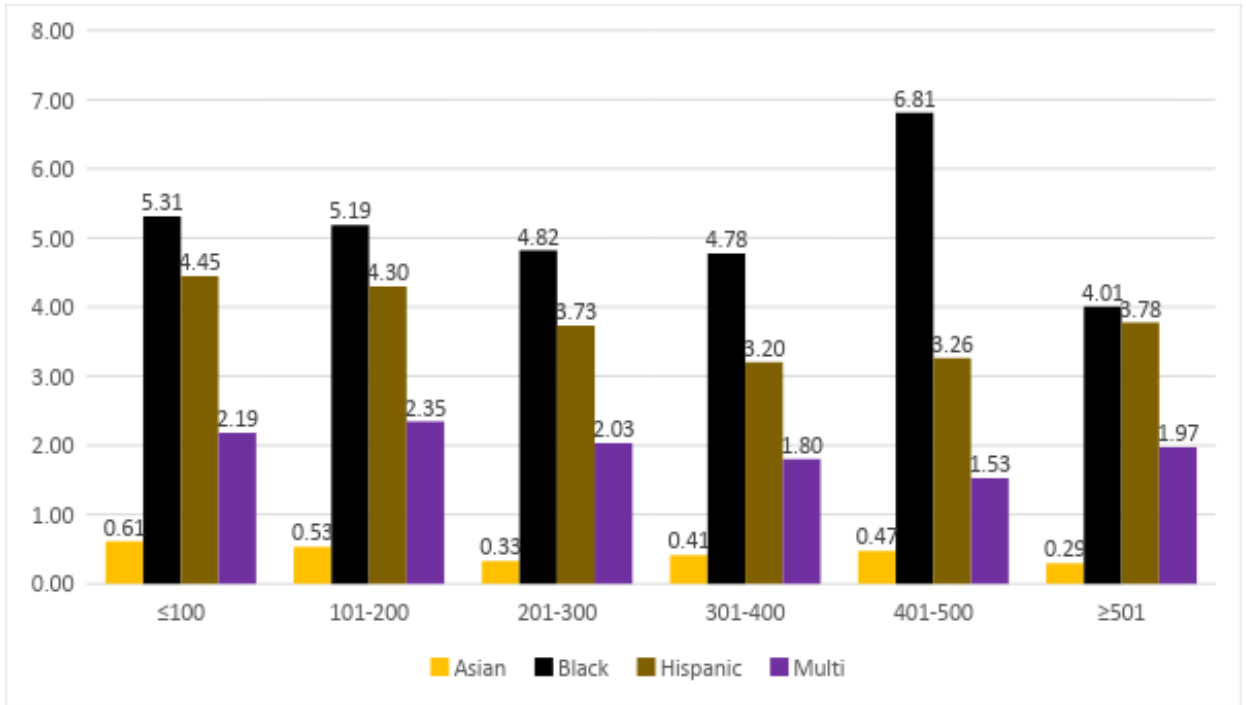


Figure 8. Shows effect-size estimates by race

Table 14 displays results for race, school size, and reading achievement.

RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Table 14.

*Estimated Effect-Size Measures on Race by School Size and Reading Achievement*

Size	Asian	Black	Hispanic	Multi
≤ 100	0.51	4.68	4.19	1.99
101 - 200	0.45	4.57	4.05	2.31
201 - 300	0.30	4.32	3.68	1.98
301 - 400	0.34	4.37	3.15	1.79
401 - 500	0.42	5.88	3.15	1.44
≥ 501	0.27	3.73	3.58	1.91

The results indicate a similar trend exists for reading achievement compared to overall achievement. The results are graphically represented in Figure 9.



RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

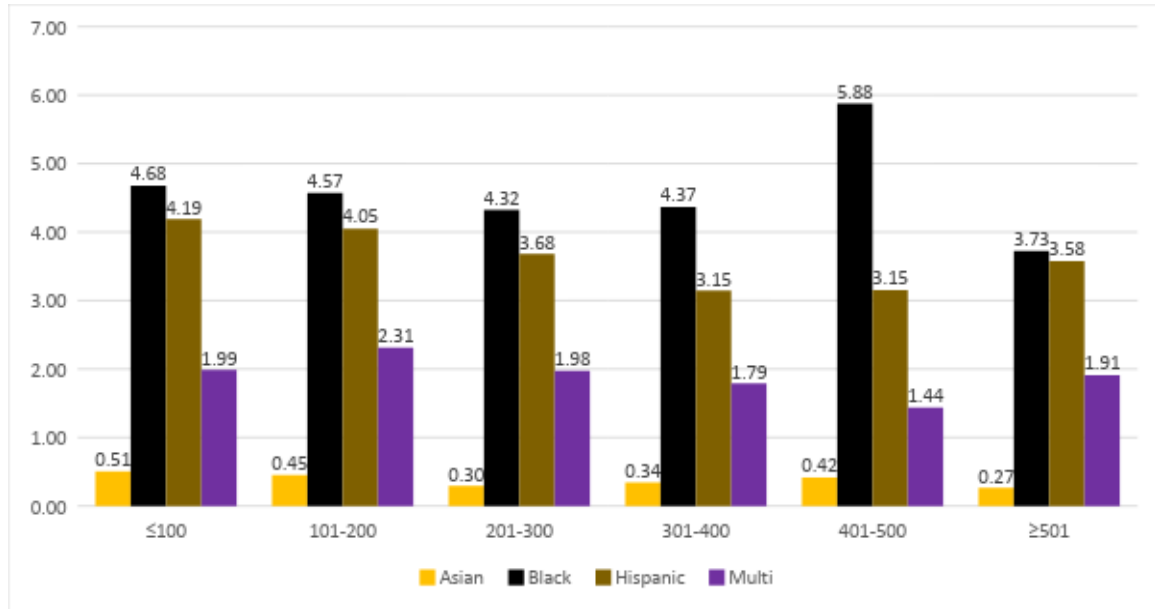


Figure 9. Shows the estimated effect-size estimates by race and reading achievement. Achievement gaps are larger when examining math achievement. These results are presented in Table 15.

Table 15.

*Estimated Effect-Size Estimates on Race by School Size and Math Achievement.*

Size	Asian	Black	Hispanic	Multi
≤ 100	0.51	6.84	5.02	2.51
101 - 200	0.45	6.60	4.74	2.40
201 - 300	0.30	5.68	3.79	2.10
301 - 400	0.34	5.46	3.26	1.81
401 - 500	0.42	9.28	3.40	1.61
≥ 501	0.27	4.45	4.06	2.04

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

The results indicate a general, downward trend in achievement gaps as schools get larger. However, a few exceptions exist. Black students in schools with 401 - 500 students demonstrate the highest gap; Hispanic students show the smallest gap in schools with 301 - 400 students; Multi-Racial students show the lowest gap in schools with 401 - 500 students. The results are graphically displayed in Figure 10.

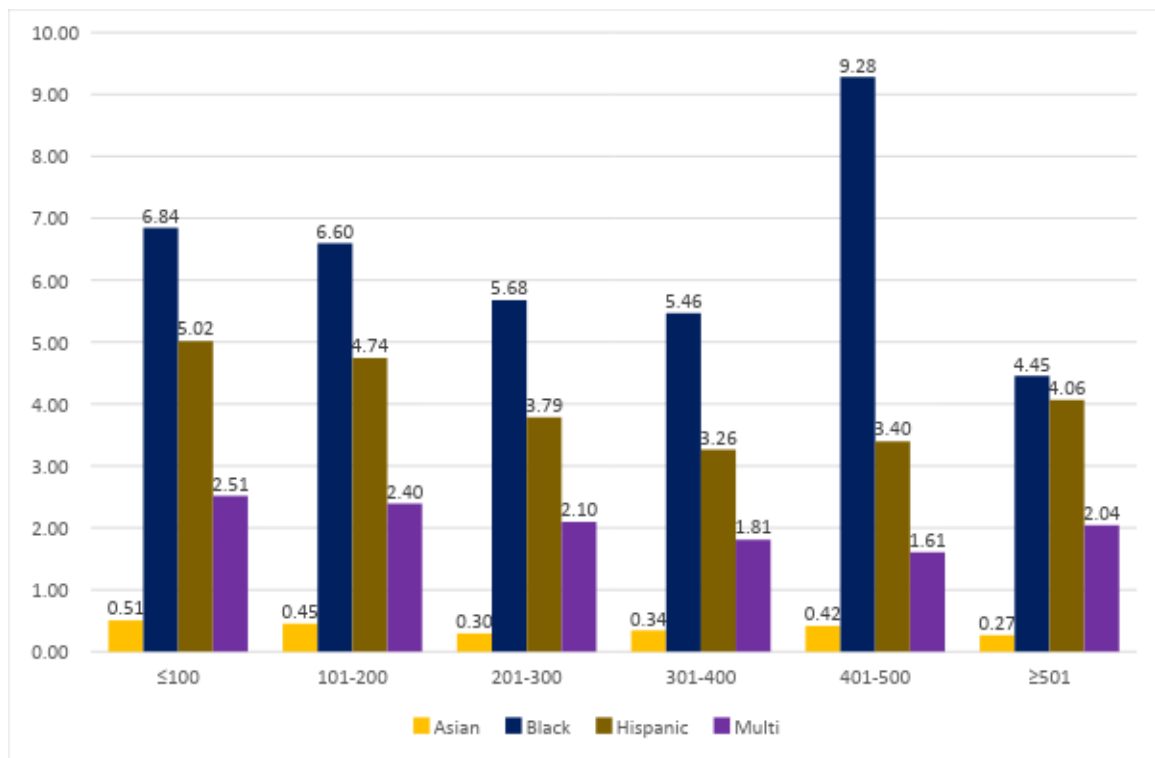


Figure 10. Shows the estimated effect-size estimates by race and math achievement

### Summary

The findings of the current study are similar to well-established trends regarding student achievement. In particular, the results showed that Pennsylvania students who receive free and reduce lunch, belong to a minority, or have a disability demonstrate a substantial achievement gap on standardized tests compared to their student peers. The

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

data also represent a confirmation and contradiction of established research concerning the subject. Specifically, PA students who are economically disadvantaged or suffer from a disability show higher achievement in smaller schools than their counterparts in larger schools. Students belonging to a minority subgroup tend to perform better in the largest schools and show the widest achievement gap in smaller schools. Chapter 5 will expound on these results, discuss the context and implication of the findings, and offer suggestions for future research.

Chapter 5

**Discussion and Conclusions**

The final chapter of this study is divided into five sections: summary, context of findings, implication of findings, limitations, and future directions. The summary section will discuss the overall findings of the study and whether or not the findings supported the research questions. The context of the findings' section will address the meaning of the results as well as reintroduce concepts from the literature review and discuss how the findings relate to previous literature and similar studies. The implications' section elaborates on theoretical frameworks, variables, alternative explanations of results, and educational implications. The limitations' section will address issues related to the studies design or shortcomings illuminated during the data analysis. The future findings' section will suggest potential research directions and other extensions of the current study.

**Summary of Findings**

The purpose of the study was to determine whether a relationship existed between school size and student achievement and whether the achievement of students categorized as economically disadvantaged, receive special education service, or belong to a minority status is affected by the size of the school. When using White, male students as the benchmark group, the findings indicated that all other students were twice as likely to not demonstrate proficiency on PSSA exams in both reading and math. When considering school size, the difference in pass rates was not practically significant. Additionally, the results showed that no practically significant difference existed between grade level

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

performance. However, when specifically breaking down subgroup categories, differences in performances existed based on school size and across grade levels.

Female students were more likely to score proficient on PSSA exams than male students. In fact, female students were over 25% more likely to be proficient on exams than male students. In reading, specifically, the female pass rate was over 50% higher than male students, but pass rates were nearly identical in math. The pass rates for female students was not significantly different based on school size.

When removing gender as a variable, students belonging to historically underperforming subgroups, including special education students, those receiving free and reduced lunch, and minority students, demonstrated significantly lower proficiency rates than White students. In fact, White students were at least four-times as likely to pass PSSA exams than members of any subgroup. Students receiving free and reduced lunches were the least likely to pass of the subgroups, followed by students belonging to a minority, and then, students with an IEP.

When looking at subgroups, more specifically students with IEPs had varying results across school size. Although students with IEPs had significantly lower pass rates than non-IEP students, they had the highest pass rates in the smallest schools (less than 100 students). The next highest pass rates were found in medium-sized schools (301-400 students) followed closely by smaller schools with 101-200 students and 201-300 students. Consequently, students in the largest schools, those with 401-500 students and 501+ students, had the lowest pass rates for students with IEPs. This trend was similar when considering test subjects as well. In reading, the highest pass rates for students with IEPs belong to the smallest school and the lowest pass rates belong to one of the

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

larger categories (401-500 students). More dramatically, in math, students in the smallest schools, less than 300 students, were nearly twice as likely to pass standardized exams than those students in the largest schools (501+ students).

Similar findings exist in regard to Socioeconomic status. When considering SES, the smallest three categories of schools (300 students or less) had the highest proficiency rates compared to the largest schools (400 or more students). The highest pass rates occurred in schools with 101-200 students, and the lowest pass rates were demonstrated in the largest category (501+ students). When considering specific tested subject, similar results existed. Students receiving free and reduced lunch had the highest reading proficiency rates in the smallest schools (200 or less students), and the largest schools (400+ students) had the lowest rates. In math, the smallest schools (200 or less students) had the highest proficiency rates, and the largest schools (501+ students) had the lowest pass rates.

When considering race overall, any student belonging to a minority subgroup, the results differed from other subgroups. Specifically, when examining race, students in the smallest schools had some of the poorest proficiency rates compared to some of the larger schools. In fact, students in the largest category (600+ students) scored proficient at nearly twice the rate of the smallest schools (less than 100 students).

The findings showed disparate results when investigating specific racial categories. Asian students, for example, fared better on standardized tests than their White counterparts. Asian students are nearly twice as likely to score proficient on PSSA exams. Furthermore, Asian students show the highest proficiency rates in the largest schools compared to the smallest schools. In fact, the Asian students are three-times

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

more likely to score proficient in the largest schools (501+ students) than the smallest school (less than 100 students).

Multi-racial students have the highest proficiency rates of the other racial subgroups, but multi-racial students are still half as likely to pass exams as White students. Although not practically significant, multi-racial students have the highest pass rates in the largest half of schools (300+ students) compared to the lower half (less than 300 students).

Hispanic students have the next highest proficiency rates but are nearly four-times less likely to pass exams than White students. Similar to multi-racial students, Hispanic students have the highest pass rates in larger schools (300+ students) compared to smaller schools (less than 300 students).

Black students are at least four-times less likely to score proficient as White students, but are as much as six-times less likely to pass the exams depending on the size of the school. Specifically, Black students are most likely to pass standardized exams in the largest schools (501+ students). Other than schools with 401-500 students, Black students are the least likely to pass exams in the smallest schools (less than 200 students).

When considering subject-specific results, similar trends remained. In reading, Asian students continued to score at higher proficiency rates than White students, with the highest rates in the largest schools and lowest rates in the smallest schools. Multi-racial students continued to be less likely than White students to score proficient in reading, but that discrepancy remains consistent regardless of school size. Hispanic students still have lower pass rates in reading than their White counterparts but have a narrower gap in mid-size schools compared to smaller schools. Black student scores still

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

lag behind White students' in reading, and those rates have some of the largest gaps in the smaller schools compared to the largest schools.

In math, Asian students continued outperform White students, and the higher achievement rates were most evident in the largest schools. Hispanic student scores continued to demonstrate a gap with White students, but the gap was similar across school-size categories. Hispanic student scores, although significantly lower than White students', showed the closest gap in mid-sized schools and the largest gap in the smaller school contingent. White students were almost 10-times more likely to pass math exams in schools with 401-500 students but were only four-times as likely in the largest schools. The gap was not its widest in the smallest schools, but it's widest in smaller schools than larger.

Overall, portions of the findings supported the research questions, and other portions did not. For instance, students who have IEPs are more likely to score proficient on PSSA exams in smaller schools than larger schools. Students who receive free and reduced lunch also showed a higher tendency to score proficient in smaller schools than larger schools. However, students belonging to a minority were actually more likely to score proficient in larger schools than smaller schools. With Black students, outliers appeared to exist in schools containing 401-500 students. The achievement gap was largest in that group, although schools with 501+ students had the smallest achievement gap.

### **Context of Findings**

As anticipated, students belonging to historically underperforming subgroups have significantly lower proficiency rates than White students. In fact, White students



## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

are twice as likely to pass PSSA exams than members of subgroups. When dividing research findings by subgroups and school size, various trends emerged within the results. The findings suggested that disability and SES subgroups have higher proficiency rates in smaller schools compared to larger schools, and members of racial subgroups tended to have higher proficiency rates in larger schools than smaller schools.

The findings in this study often mimicked the results of previous research, but portions of the findings also contradicted previous research assertions. While looking at each subgroup category specifically, this study's findings will be compared and contrasted with aspects of the literature review, specifically discussing results, procedures, and variables.

Just as Silvernail et al. (2014) stated that poverty is the single best predictor of student achievement, the results of the 2015-2018 PSSA exams showed a strong correlation between free- and reduced-lunch status and student achievement. Researchers such as Howley and Bickel (2000) suggested the negative effects of poverty can be exacerbated as schools become larger, and Howley and Howley (2004) showed that poorer students show higher achievement in smaller schools compared to their economically disadvantaged counterparts in larger schools.

The results of this study appeared to mirror those assertions that students receiving free or reduced lunches demonstrated higher proficiency rates on PSSA exams in smaller schools. In the smaller school categories, economically disadvantaged students are 3.93 times less likely to score proficient on standardized exams compared to non-disadvantaged students. In the largest school categories, that number rises to 4.9 times less likely. In other words, students in smaller schools are 20% more likely to pass

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

state exams than students in the largest schools. Furthermore, that number consistently rises as the category of school size increases.

Riggen (2013) stated that students with IEPs in Illinois are nearly five-times more likely to score proficient on standardized exams compared to students without disabilities. The results in this study are more dramatic, where students in Pennsylvania can be as much as nine-times more likely to pass PSSA exams than students with a disability. Gershenson and Langbein (2015) claimed that students were particularly sensitive to increases in school size, and data from this study seem to support this claim. In the current investigation, achievement gap was the narrowest at the smallest school size and was the widest in the largest schools, those with 500 or more students. This trend was more apparent when investigating mathematics in particular, where students in smaller schools were twice as likely to score proficient on PSSA exams.

Although many researchers have claimed students with disabilities have higher achievement in smaller schools, they very rarely agree on the reason. Kirby (2017) contended that smaller schools have unique opportunities to implement inclusionary techniques, and mainstream time in general education can lead to higher achievement. Marks, Kurth, and Bartz (2014) contended the most important feature of inclusive settings is frequently school size, and larger school districts show the lowest percentage of mainstreaming. Although this study does not attempt to offer potential reason for the increase in the achievement gap, it does support the notion that special education students fare better in smaller schools.

In regard to race, the current investigation helped confirm a national trend that a substantial achievement gap exists between White students and minority students. In

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Pennsylvania, White students are five-times more likely to be proficient than Black students, four-times more likely than Hispanic students, and two-times more likely than multiracial students. According to Webb and Thomas (2015), Black and Hispanic students demonstrated nearly a 20-30% gap between White students on national exams. For PSSA exams, that gap is even wider, where Black students showed a 35% gap, and Hispanic students demonstrated a 30% gap. Webb and Thomas (2015) suggested that a 20-point gap can equate to approximately two grade levels. In turn, this study may be highlighting a larger, three-grade-level gap between Black and White students.

Achilles (2012) indicated that student achievement was significantly higher for Black students in smaller schools. Achilles (2012) did not utilize standardized test results as the basis of the claims, but he utilized grade retentions, dropout rates, and college entrance exams. However, the results of this study appear to indicate that Black students perform higher on state standardized tests in larger schools than smaller schools. In fact, Black students in Pennsylvania are 20% more likely to score proficient on state assessment exams in the largest set of schools compared to the smallest set of schools. In other words, White students are four-times more likely to pass PSSA exams in the largest schools, but that number increases to 5.3 times in the smallest schools.

Howley and Howley (2004) found that standardized test scores were the highest in mid-sized schools, but the achievement gap was most narrow in smaller schools. Furthermore, Howley and Howley (2004) demonstrated the lowest achievement and widest gap existed in the large school category. Again, the current study suggested that Black students have the narrowest gap in the largest school group and widest in the smallest school group. However, the data suggest the gap in mid-size schools is much

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

closer to the gap in larger schools. For instance, the narrowest gap in the largest schools is 4.01 and in the mid-size schools is 4.8.

Zoda et al. (2011) found that Hispanic student achievement in reading, although significantly behind White students', was the highest in larger schools than in smaller schools, and, although not statistically significant, math achievement was higher on average in larger schools versus smaller schools. This study seemed to develop similar results. Although the narrowest gap existed in mid- to large-size schools, the largest gap existed in the smaller tier of schools.

### **Implications of the Findings**

Areas of this study are consistent with a wide range of research and align with many theories of student achievement and achievement gaps, and other areas of the study contradict aspects of theoretical frameworks. Regardless, the study provides Pennsylvania educators a foundation of information regarding core standard achievement, school size, and the historically underperforming achievement gap.

Overall, and regardless of school size, the results confirm a concerning achievement gap exists across many subgroup categories in the state. The findings show a relationship exists between school size and achievement. For students who are economically disadvantaged and those with disabilities, the data suggest they will fare better on PSSA exams in smaller schools. Students belonging to a minority subgroup appear to perform the best in the largest tiered schools and demonstrate the widest achievement gaps in mid- to larger-sized schools and wider gaps in the smaller tiered schools.

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

The value of this investigation is that it analyzes variables after the shift to PA Core Standards and utilizes a school-size independent variable. However, it is imperative to consider that any relationship to school size and achievement may be attributed to many potential confounding variables. For example, IEP students have the narrowest achievement gap in smaller schools, but this relationship may be attributed to inclusionary advantages, communication and continuity between the school and family, the potential familiar relationship between staff and students, or other theoretical assertions. Students receiving free or reduced lunches perform more closely to non-disadvantaged peers in smaller schools, but the specific aspects experienced by these students in small versus larger schools cannot solely be determined from these findings. Conversely, minority students are performing well in the largest tier of schools, so attempting to identify the particular aspects of that environment would be the next practical step to investigate. The largest schools may have a more vast and more diverse curriculum; they may have more support services for students and families; they may have more fiscal resources, and/or many other amenities. Regardless, Pennsylvania legislators and educational leaders should consider what characteristics of small and large schools appear most beneficial and take steps to support schools to adopt the strategies.

### **Limitations**

Although a need exists to evaluate PA schools, the external validity of this study would be suspect outside of the sample. This study does not include high school students and their performance on standardized tests, the Keystone Exams. Using the free- and reduced-lunch rates is a fairly vast metric, and it does not necessarily offer potentially valuable details. For example, families may qualify for free and reduced lunches with

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

annual salaries 100% to 180% above the poverty line. Determining the differences between the poorest students and those who barely qualify may be useful, but impossible with this structure. The simplistic synthesizing analysis does not allow for any school culture analysis or gathering any qualitative data. As with any linear analysis, the PSSA data would be sensitive to outliers.

Although the data may suggest a relationship exists between school size and student achievement, merely identifying the size of the school as the reason for higher or lower achievement would be premature and overly simplistic. As the literature review suggested, a myriad of factors have an influence on student achievement and many factors, including school size, may create the existence of those factors and the influence they have on student achievement. Issues including but not limited to school climate, teacher attitude and beliefs, inclusionary methods, and others may have a substantial impact on student achievement. Considering this dynamic, the study may illuminate trends in regard to the relationship of school size and student achievement and may highlight the potential existence of external factors affecting achievement, but the scope of the study will not permit assertions on external factors nor will it allow for the elaboration on those factors.

The current study did not consider various aspects of school that may affect student achievement. For instance, NCES (2015) discussed a school's racial density or the percentage of minority students in the school. The researchers determined that schools with higher density tended to demonstrate lower achievement on standardized assessments. For example, Black students who attended schools with large proportions

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

of Black students scored more poorly than Black students in schools with lower percentages of Black students. The current study did not account for school density.

The current study did not account for combinations of composition factors. To illustrate, the findings did not delineate if a student belonged to multiple or all subgroup categories. A student could theoretically belong to all three of the subgroup factors, such as a Hispanic student who receives free lunch and has an IEP. If the student in this example did not score proficient, the student would count as a non-proficient student overall and for each subgroup category.

All the relationships investigated in the study produced a significant p-value, but the large sample size may be a cause. This study is heavily reliant on the PDE and the Data Recognition Corporation's ability to limit test measurement error within the assessment. The information utilized to determine SES, IEP rates, and minority membership is based on a self-report system in school districts and is subject to their accuracy.

Although the Pennsylvania Department of Education considers proficiency on standardized tests as a measure of success, many other indicators of student learning and competency exist. While there are a plethora of metrics to judge success, a more stable measure of student productivity is their ability to become a productive and contributing citizen.

### **Future Research**

The findings that minority students have wider gaps in smaller schools is not only contradictory to the hypothesis but is contradictory to some previous research findings. This dynamic would require further investigation in specific areas. In the Council for

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

Great City Schools (2012), researchers claimed that minority students tend to be in larger schools: 93 more students on average. Looking into the trend of school size and minority enrollment would be beneficial for further understanding. As suggested in the limitation section, school composition was not utilized in the current study and likely has an impact on student achievement. Looking at school size and composition together would be warranted.

In another reference from the Council for Great City Schools (2012), the authors suggested that minority students often have larger student-to-teacher ratios, 1:18 compared to 1:16. Further studies into student/teacher ratios and their impact on achievement would be a prudent next step. Tienken and Achilles (2009) suggested that class size is much different from student/teacher ratio because actually counting students in the classroom is the only accurate method to determining class size. Tienken and Achilles (2009) suggested that small schools do not necessarily have small class sizes, and conversely, large schools do not necessarily have large class sizes. A post-common core Pennsylvania class size investigation would be worthy of future research.

The NCES (2015) suggested that Black male students demonstrate the largest achievement gap with their White peers. Looking at Black male achievement and school size would be a relevant next step. Furthermore, looking at any subgroup performance separated by gender would be worth investigating. For example, researchers could compare Hispanic male versus female performance and evaluate changes in the achievement gap. The same could be done for SES and disabilities.

In the NCES (2015) racial density study, researchers indicated that students in high-density schools, those with large percentages of minority students were less likely to



## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

score proficient on standardized exams. Moreover, minority students were much more likely to attend high-density schools than White students. Since racial density is intensified in Pennsylvania compared to national rates, a study evaluating the relationship between density rates and student achievement would be warranted.

Abbott et al. (2002) claimed that students in smaller schools are more likely to feel a sense of community and maintain positive social relationships in smaller schools compared to larger schools, and students who develop these and other similar characteristics are a part of a healthy school climate. Jones and Shindler (2003) so strongly supported the concept that they claimed school climate is one of the most predictive factors in any school's capacity to improve student performance.

Studying the relationship of school climate and student achievement would be an excellent topic for future research. Pennsylvania schools already have a state-developed, standardized, school-climate survey that can be paired with student achievement results to investigate any potential relationship.

Furthermore with Abbott et al. (2002), future qualitative research regarding perceptual data and student achievement would be pertinent and applicable. Researchers could not only evaluate the relationship of student perceptions of the schools but could also investigate parental perceptions. Goldkind and Farmer (2013), using a mixed methods design, studied parental perceptions of their children's schools and found parents in smaller schools believed the educational environment was more favorable. The parents in smaller schools believed the educational environment was safer, more respectful, and offered more opportunities for parental involvement. Future researchers could investigate a potential qualitative correlation and student achievement.

## **Conclusion**

School size and student achievement have been vigorously studied over the past few decades, with many researchers establishing a relationship between the two variables. Many have claimed that student achievement is higher in smaller schools, and others have determined no relationship exists. Some researchers have suggested that smaller schools help historically underperforming students, such as those belonging to minorities' subgroups, those who are economically disadvantaged, those with disabilities, and others. In this study, the findings indicated that students in smaller schools who receive free and reduced lunches and those with disabilities have higher achievement than those in larger schools. The findings also indicated that students belonging to a minority population, Black, Hispanic, Asian, or Multi-Racial have higher achievement in the largest schools and lowest achievement in the smallest schools. Regardless of the relationship, many confounding variables may exist in school environments that have an effect or combined effect on the achievement of students; therefore, establishing a clear causal relationship between school size and student achievement would be irresponsible. However, the findings are adequately compelling to spur further research regarding school size and other associated variables such as class size, school climate, and racial/economic diversity. Regardless of the size of our schools, educational leaders can take steps to optimize the educational environment and identify various factors that impact student performance.

References

- Abbott, M. L., Joireman, J., & Stroh, H. R. (2002). *The influence of district size, school size and socioeconomic status on student achievement in Washington: A replication study using hierarchical linear modeling*. (Technical Report). Retrieved from <http://files.eric.ed.gov/fulltext/ED470668.pdf>
- Achieve. (2005). *Measuring up: A report on assessment anchors and tests in reading and mathematics for PA*. Retrieved from Achieve website: <https://doi.org/10.1186/1472-6955-12>
- Achilles, C. M. (2012). Class-size policy: The star experiment and related class-size studies. *NCPEA Policy Brief, 1*(2), 10.
- Alspaugh, J. W., & Gao, R. (2003). School size as a factor in elementary school achievement. Retrieved from ERIC database. (ED475062)
- Bohrnstedt, G., Kitmitto, S., Ogut, B., Sherman, D., and Chan, D. (2015). School Composition and the Black–White Achievement Gap (NCES 2015-018). U.S. Department of Education, Washington, DC: National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubsearch>
- Bol, L., & Berry III, R. Q. (2005). Secondary mathematics teachers' perceptions of the achievement gap. *High School Journal, 88*, 32–45. <https://doi.org/10.1353/hsj.2005.0007>
- Bowne, J. B., Magnuson, K. A., Schindler, H. S., Duncan, G. J., & Yoshikawa, H. (2017). A meta-analysis of class sizes and ratios in early childhood education programs: Are thresholds of quality associated with greater impacts on cognitive

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

- achievement and socioemotional outcomes? *Educational Evaluation and Policy Analysis*, 39(3), 407–428.
- Choi, K., Goldschmidt, P., & Yamashiro, K. (2006). Exploring models of school performance : From theory to practice. *National Center for Research on Evaluation Standards, and Student Testing*, 1–31. Retrieved from <https://doi.org/10.1016/j.oraloncology.2008.05.008>
- Cohen, J., Pickeral, T., & McCloskey, M. (2009). Assessing school climate. *Educational Leadership*, 74(8), 45–48.
- Coladarci, T. (2006). School size, student achievement, and the “power rating” of poverty: Substantive finding or statistical artifact? *Education Policy Analysis Archives*, 14(28), 1-26.
- Coulson, A. J. (2007). School district consolidation, size and spending: An evaluation. Retrieved from <https://files.eric.ed.gov/fulltext/ED520065.pdf>
- Council for Great City Schools. (2012). Beating the odds: Analysis of student performance of state assessments. Retrieved from <https://www.cgcs.org/cms/lib/DC00001581/Centricity/Domain/87/FINALBTO12%2012.pdf>
- Darling-Hammond, L. (2015). Want to close the achievement gap? *American Federation of Teachers*, 1–6. Retrieved from <http://www.aft.org/ae/winter2014-2015/darling-hammond>
- Data Recognition Corporation. (1978). Retrieved from <https://www.datarecognitioncorp.com/Pages/default.aspx>

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

- Frisbie, D. A. (2005). Measurement 101: Some fundamentals revisited. *Educational Measurement: Issues and Practice*, 24, 21–28. Retrieved from <https://doi.org/10.1111/j.1745-3992.2005.00016.x>
- Gershenson, S., & Langbein, L. (2015). The effect of primary school size on academic achievement. *Educational Evaluation and Policy Analysis*, 37, 135S–155S. Retrieved from <https://doi.org/10.3102/0162373715576075>
- Goldkind, L., & Farmer, L. G. (2013). The enduring influence of school size and school climate on parents' engagement in the school community. *School Community Journal*, 23(1), 223–244. Retrieved from <http://search.proquest.com/docview/1373086990?accountid=14609>
- Governor's Press Release. (2017). Governor Wolf announces more reductions in standardized testing for Pennsylvania students, teachers. Retrieved from <https://www.governor.pa.gov/governor-wolf-announces-additional-standardized-testing-reductions-pennsylvania-students-teachers/>
- Howley, C., & Bickel, R. (2000). When it comes to schooling...small works: School size, poverty, and student achievement. *Journal of Poverty*, 6(3), 89–108.
- Howley, C., Johnson, J., & Petrie, J. (2011). *Consolidation of schools and districts: What the research says and what it means*. Boulder, CO: National Education Policy Center. Retrieved from <http://nepc.colorado.edu/publication/consolidation-schools-districts>
- Howley, C. B., & Howley, A. A. (2004). School size and the influence of socioeconomic status on student achievement: Confronting the threat of size bias in national data

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

- sets. *Education Policy Analysis Archives*, 12(52), 1–35. Retrieved from <http://epaa.asu.edu/ojs/article/view/227>
- Iatarola, P., Conger, D., & Long, M. C. (2011). Determinants of high schools' advanced course offerings. *Educational Evaluation and Policy Analysis*, 33, 340–359. Retrieved from <https://doi.org/10.3102/0162373711398124>
- Januszka, C., & Dixon-Krauss, L. (2008). Class size: A battle between accountability and quality instruction. *Childhood Education*, 84, 167-170. Retrieved from <https://doi.org/10.1080/00094056.2008.10523000>
- Jepsen, C., & Rivkin, S. G. (2013). Class reduction and student achievement: The potential tradeoff between teacher quality and class size. *Journal of Human Resources*, 53, 1689–1699. Retrieved from <https://doi.org/10.1017/CBO9781107415324.004>
- Johnson, J. (2006). *More doesn't mean better*. Arlington, VA: Rural School and Community Trust.
- Joint State Government Commission. (2017). *School districts in PA: Ways to work together. 2015 House Resolution 910 Staff Study*. Retrieved from <http://jsg.legis.state.pa.us>
- Jones, A., & Shindler, J. (2016). Exploring the school climate—student achievement connection: Making sense of why the first precedes the second. *Educational Leadership and Administration: Teaching and Program Development*, 27(2003), 35–51.

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

- Jones, K. R., & Ezeife, A. N. (2011). School size as a factor in the academic achievement of elementary school students. *Psychology*, 2, 859–868. Retrieved from <https://doi.org/10.4236/psych.2011.28131>
- Jordan, A. H., & Lovett, B. J. (2007). Stereotype threat and test performance: A primer for school psychologists. *Journal of School Psychology*, 45, 45–59. Retrieved from <https://doi.org/10.1016/j.jsp.2006.09.003>
- Kirby, M. (2017). Implicit assumptions in special education policy: Promoting full inclusion for students with learning disabilities. *Child & Youth Care Forum*, 46, 175–191. Retrieved from <https://doi.org/10.1007/s10566-016-9382-x>
- Konstantopoulos, S. (2008). Do small classes reduce the achievement gap between low and high achievers? Evidence from project star. *The Elementary School Journal*, 108, 275–291. Retrieved from <https://doi.org/10.1086/528972>
- Lane, S. (2014). Validity evidence based on testing consequences. *Psicothema*, 26, 127–135. Retrieved from <https://doi.org/10.7334/psicothema2013.258>
- Leahy, S. (2006). A survey of selected teacher opinions to the effects of class size on student achievement among middle school students. Retrieved from ERIC database (ED494718)
- Lee, V. E., & Loeb, S. (2000). School size in Chicago elementary schools: Effects on teachers' attitudes and students' achievement. *American Educational Research Journal*, 37(1), 3-31.
- Lee, V. E., & Smith, J. B. (1996). *High school size: Which works best, and for whom?* Thousand Oaks, CA: Sage Journals.

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

- Lockwood, J. R., & McCaffrey, D. F. (2007). Controlling for individual heterogeneity in longitudinal models, with applications to student achievement. *Electronic Journal of Statistics, 1*, 223–253. Retrieved from <https://doi.org/10.1214/07-EJS057>
- Mann, M. J., Maxwell, G. M., & Holland, G. (2013). Differences in middle school science achievement by school district size. *Journal of Instructional Pedagogies, 12*, 1–6. Retrieved from <http://jupapadoc.startlogic.com/manuscripts/131599.pdf>
- Marks, S. U., Kurth, J. A., & Bartz, J. M. (2014). Exploring the landscape of inclusion: Profiles of inclusive versus segregated school districts in the United States. *The Journal of the International Association of Special Education, 15*(2), 74–82.
- Pennsylvania Department of Education. (2018, November). *PVAAS methodologies: Measuring growth & projecting achievement*. Harrisburg, PA: Author.
- Pennsylvania Department of Education. (2017). *2017 PA state system of assessment technical report*. Harrisburg, PA: Author.
- Pennsylvania School Boards Association. (2009). *Merger / consolidation of school districts: Does it save money and improve student achievement?* Mechanicsburg, PA: Author.
- Peterson, E. R., Rubie-Davies, C., Osborne, D., & Sibley, C. (2016). Teachers' explicit expectations and implicit prejudiced attitudes to educational achievement: Relations with student achievement and the ethnic achievement gap. *Learning and Instruction, 42*, 123–140. Retrieved from <https://doi.org/10.1016/j.learninstruc.2016.01.010>
- Reardon, S. F. (2013). The widening income achievement gap. *Educational Leadership, 70*(8), 10. Retrieved from



## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

<http://www.systems.wsu.edu/scripts/wsual1.pl?url=http://search.ebscohost.com/login.aspx?direct=true&db=ulh&AN=87529513&site=ehost-live>

Reid, S. (2002). *The achievement gap: How minority students are faring in North Carolina public schools*. North Carolina Justice and Community Development Center, Raleigh, NC.

Riggen, V. A. (2013). *School size and student achievement*. (Unpublished doctoral dissertation). Indiana State University, Terre Haute, IN.

Rural School and Community Trust. (2002). *Small works in Arkansas: How poverty and the size of schools and school district affect school performance in Arkansas*. Alexandria, VA: author.

Scheerens, J., Hendriks, M., & Luyten, H. (2014). School size effects revisited. *Springer Briefs in Education*, 7–40. Retrieved from <https://doi.org/10.1007/978-3-319-06814-5>

Silvernail, D. L, Sloan, J. E., Paul, C. R., Johnson, A. F., & Stump, E. K. (2014). *The relationships between school poverty and student achievement in Maine*. Gorham, ME: Maine Education Policy Research Institute. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED561363&site=ehost-live>

Standard and Poor's School Evaluation Services. (2008). Retrieved from (<http://static.usnews.com/documents/best-highschools/methodology>

Statistical Analysis System Educational Value-Added Assessment System. (2016). Retrieved from [https://www.sas.com/en\\_us/software/evaas.html](https://www.sas.com/en_us/software/evaas.html)

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

- The Standards for Educational Psychological Testing. (2014). Retrieved from <https://www.apa.org/science/programs/testing/standards.aspx>
- Tienken, C. H., & Achilles, C. M. (2009). Relationship between class size and students' opportunity to learn writing in middle school. *Research in the Schools, 16*(1), 13–24.
- Uerling, D., & Dlugosh, L. (1999). *Selected indicators of a quality high school: Program offerings and student participation*. Paper presented at the annual conference on Creating the Quality Schools, Memphis, TN.
- Viadero, D. (2008). Students observed to be on task less as class size grows. *Education Week, 27*(31), 9-12.
- Wasserberg, M. J., & Rottman, A. (2016). Urban high school students' perspectives on test-centered curriculum: EBSCOhost. *American Secondary Education, 44*(3), 56–71. Retrieved from <http://web.a.ebscohost.com.proxy-s.mercer.edu/ehost/pdfviewer/pdfviewer?vid=11&sid=12175dc4-be9c-4a28-a7fd-f125df0a669c%40sessionmgr4007>
- Webb, M., Thomas, R. (2015). Teachers' perceptions of educators' and students' role in closing the achievement gap. *National Forum of Teacher Education, 25*(3), 1–9.
- White, G. W., Stepney, C. T., Hatchimonji, D. R., Moceri, D. C., Linsky, A. V., Reyes-Portillo, J. A., & Elias, M. J. (2016). The increasing impact of socioeconomics and race on standardized academic test scores across elementary, middle, and high school. *American Journal of Orthopsychiatry, 86*, 10–23. Retrieved from <https://doi.org/10.1037/ort0000122>

RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

- Wood, L., Kiperman, S., Esch, R. C., Leroux, A. J., & Truscott, S. D. (2017). Predicting dropout using student- and school-level factors: An ecological perspective. *School Psychology Quarterly, 32*, 35–49. Retrieved from <https://doi.org/10.1037/spq0000152>
- Zoda, P., Slate, J. R., & Combs, J. P. (2011). Public school size and Hispanic student achievement in Texas: A 5-year analysis. *Educational Research for Policy and Practice, 10*, 171–188. Retrieved from <https://doi.org/10.1007/s10671-011-9103-z>

## Appendix A

### **Inclusive Assessment Population**

The PSSA target population includes all students at the assessed grades attending Commonwealth schools. For state, district, and school accountability purposes, the target population includes all students except those who will participate in accountability through an alternate assessment.

### **Precisely Defined Constructs**

An important function of well-designed assessments is that they actually measure what they are intended to measure. The PA Assessment Anchors and Eligible Content provided clear descriptions of the constructs to be measured by the PSSA at the assessed grade levels. Universally designed assessments must remove all non-construct-oriented cognitive, sensory, emotional, and physical barriers.

### **Accessible, Non-biased Items**

DRC conducted both internal and external reviews of items and test specifications to ensure that they did not create barriers because of lack of sensitivity to disability, culture, or other subgroups. Items and test specifications were developed by a team of individuals who understand the varied characteristics of items that might create difficulties for any group of students. Accessibility is incorporated as a primary dimension of test specifications, so accessibility was woven into the fabric of the test rather than added after the fact. The following examples show two graphics with the same construct, example 1 being less accessible and example 2 being more accessible.

### **Amenable to Accommodations**

Even though items on universally designed assessments are accessible for most students, there are some students who continue to need accommodations. This essential element of a universally designed assessment requires that the test is compatible with accommodations and a variety of widely used adaptive equipment and assistive technology. (See the section on Assessment Accommodations later in Chapter Four.)

### **Simple, Clear, and Intuitive Instructions and Procedures**

Assessment instructions should be easy to understand, regardless of a student's experience, knowledge, language skills, or current concentration level. Questions that are posed using complex language can invalidate the test if students cannot understand how they are expected to respond to a question. To meet this guideline, directions and questions were prepared in simple, clear, and understandable language that underwent multiple reviews.

### **Maximum Readability and Comprehensibility**

A variety of guidelines exist to ensure the maximum readability and comprehensibility of a test. These features go beyond what is measured by readability formulas. Readability and comprehensibility are affected by many factors, including student background, sentence difficulty, text organization, and others. All of these features were considered as item text was developed.

Plain language is a concept now being highlighted in research on assessments. Plain language has been defined as language that is straightforward and concise. The following strategies for editing text to produce plain language were used during the editing process of the new PSSA items:

- Reduction of excessive length

  - Use of common words

  - Avoidance of ambiguous words

  - Avoidance of irregularly spelled words

- Avoidance of proper names

- Avoidance of inconsistent naming and graphic conventions

- Avoidance of unclear signals about how to direct attention

### **Maximum Legibility**

Legibility is the physical appearance of text, the way that the shapes of letters and numbers enable people to read text easily. Bias can result when tests contain physical features that interfere with a student's focus on or understanding of the constructs that test items are intended to assess. A style guide developed and updated annually (DRC, 2004–2013) was utilized, with PDE approval, which included dimensions of style consistent with universal design

## Appendix B

**Items measure what they are intended to measure.** Item-writing training included ensuring that writers and reviewers had a clear understanding of PA’s Core Standards (ELA and mathematics) or Academic Standards (science) and the Assessment Anchors. During all phases of test development, items were presented with content-standard information to ensure that each item reflected the intended Assessment Anchor. Careful consideration of the content standards was important in determining which skills involved in responding to an item were extraneous and which were relevant to what was being tested. In certain types of items, an additional skill is necessary, such as the mathematics test, which requires the student to read.

**Items respect the diversity of the assessment population.** To develop items that avoid content that might unfairly advantage or disadvantage any student subgroup, item-writers, test-developers, and reviewers were trained to write and review items for issues of bias, fairness, and sensitivity. Training also included an awareness of, and sensitivity to, issues of cultural and regional diversity.

**Items have a clear format for text.** Decisions about how items are presented to students must allow for maximum readability for all students. Appropriate fonts and point sizes were employed with minimal use of italics, which is far less legible and is read considerably more slowly than standard typeface. Captions, footnotes, keys, and legends were at least a 12-point size **1**. Legibility was enhanced by sufficient spacing between letters, words, and lines. Blank space around paragraphs and between columns and staggered right margins were used.

**Stimuli and items have clear pictures and graphics.** When pictures and graphics were used, they were designed to provide essential information in a clear and uncluttered manner. Illustrations were placed directly next to the information to which they referred, and labels were used where possible. Sufficient contrast between background and text, with minimal use of shading, increased readability for students with visual impairments. Color was not used to convey important information.

**Items have concise and readable text.** Linguistic demands of stimuli and items can interfere with a student’s ability to demonstrate knowledge of the construct being assessed. During item-writing and review, the following guidelines were used.

- Simple, clear, commonly-used words were used whenever possible.
- Extraneous text was omitted.
- Vocabulary and sentence complexity were appropriate for the grade level being assessed.
- Technical terms and abbreviations were used only if they were related to the content being measured.
- Definitions and examples were clear and understandable.
- Idioms were avoided unless idiomatic speech was being assessed.
- The questions to be answered were clearly identifiable.

## RESOURCES, PUPIL-TYPE, OR PERSONAL ATTENTION

**Items allow changes to format without changing meaning or difficulty.** A Braille version of the PSSA was available at each assessed grade. Attention was given to using items that allow for Braille. Specific accommodations were permitted, such as signing to a student, the use of oral presentation under specified conditions, and the use of various assistive technologies. Spanish versions of the PSSA mathematics and PSSA science tests were available for use by English Language Learners who would benefit from this accommodation. In the online format, permitted accommodations included text-to-speech audio, a color overlay, contrasting text options, and American Sign Language videos.

**The test has an overall appearance that is clean and organized.** Images, pictures, and text that may not be necessary (e.g., sidebars, overlays, callout boxes, visual crowding, shading) and that could be potentially distracting to students were avoided. Also avoided were purely decorative features that did not serve a purpose. Information was organized in a left-right, top-bottom format.

October 12, 2018

Dr. Karen Larwin, Principal Investigator  
Mr. Michael Ross, Co-investigator  
Department of Counseling, School Psychology & Educational Leadership  
UNIVERSITY

RE: HSRC PROTOCOL NUMBER: 029-2019  
TITLE: Programs, Pupil-Type or Personal-Attention: The Impact of School Size of  
Student Achievement

Dear Dr. Larwin and Mr. Ross:

The Institutional Review Board has reviewed the abovementioned protocol and determined that it is exempt from full committee review based on a DHHS Category 4 exemption.

Any changes in your research activity should be promptly reported to the Institutional Review Board and may not be initiated without IRB approval except where necessary to eliminate hazard to human subjects. Any unanticipated problems involving risks to subjects should also be promptly reported to the IRB.

The IRB would like to extend its best wishes to you in the conduct of this study.

Sincerely,

Dr. Grég Dillon  
Interim Associate Vice President for Research  
Authorized Institutional Official

GD:cc

c: Dr. Jake Protivnak, Chair  
Department of Counseling, School Psychology & Educational Leadership