ETD Archive

2018

The Impact of Tracking Students in Mathematics on Middle School Student Achievement Outcomes

David P. Glasner

Follow this and additional works at: https://engagedscholarship.csuohio.edu/etdarchive

Part of the Science and Mathematics Education Commons, and the Secondary Education Commons

How does access to this work benefit you? Let us know!

THE IMPACT OF TRACKING STUDENTS IN MATHEMATICS ON MIDDLE

SCHOOL STUDENT ACHIEVEMENT OUTCOMES

DAVID P. GLASNER

Bachelor of Arts in History

University of Pennsylvania

May 2000

Master of Arts in History

University of Pennsylvania

May 2000

Master of Arts in Social Studies Education

Teachers College, Columbia University

October 2003

Submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY IN URBAN EDUCATION

at the

CLEVELAND STATE UNIVERSITY

December 2018

We hereby approve the dissertation of

DAVID P. GLASNER

Candidate for the Doctor of Philosophy in Urban Education Degree

This Dissertation has been approved for the
Office of Doctoral Studies,
College of Education and Human Services
and
CLEVELAND STATE UNIVERSITY,
College of Graduate Studies by

Dissertation Chairperson: Frederick Hampton
Counseling, Administration, Supervision, and Adult Learning

Methodologist: Adam Voight
Curriculum and Foundations

Member: Mark Freeman
Counseling, Administration, Supervision, and Adult Learning

Member: Glenda Toneff-Cotner
Curriculum and Foundations

Member: Jeffrey Snyder
Urban Studies

October 26, 2018
Candidate's Date of Defense

DEDICATION

For Elana, my partner in life, work and family. I could not have done this without you.

ACKNOWLEDGMENTS

THE IMPACT OF TRACKING STUDENTS IN MATHEMATICS ON MIDDLE SCHOOL STUDENT ACHIEVEMENT OUTCOMES

ABSTRACT

DAVID P. GLASNER

The purpose of this study was to explore whether and how tracking structures in mathematics courses at the middle school level relate to differences in achievement between white and black students. This study used propensity score matching to compare the achievement outcomes of students enrolled in advanced mathematics classes, with students of comparable ability and background enrolled in grade-level math classes. The study sample was comprised of 1,510 students.

Results from the study show that enrollment in an advanced-math course was associated with statistically significant improvement in math achievement for average-ability students. In addition, study results show that increases in student achievement associated with average-ability black student enrollment in advanced-level math courses surpass the increases in math achievement outcomes associated with average-ability white student enrollment in advanced-level math courses. These findings have important equity implications because average-ability black students opt to enroll, or are disproportionately placed, in grade-level math as compared to average-ability white students. The findings suggest that increased enrollment of average-ability black and white students in advanced-level math would lead to a reduction in the racial math achievement gap and to improved math achievement outcomes for both black and white students.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Since its inception, America has wrestled with a contradiction inherent to its existence. The country was forged with lofty aspirations of equality and democracy for all, and with a vision to create a unified nation out of a group of people that lack any type of shared national or cultural heritage (Delbanco, 2000). Yet these values were promulgated over the backs of slaves and through a history of racism that belies the ideals of this country's independence. This tension continues to trace fault lines through the American social fabric and is particularly evident in the challenges confronting the American system of public education today.

The early history of public schools in America reflects this paradox. Political leaders called for the spread of a public education system as a means to promote equality throughout the land and to bring disparate people together almost immediately after the American Revolution (Hochschild & Scovronick, 2003; Reese, 2005). For example, within three years of the writing of the Declaration of Independence, Thomas Jefferson began championing free and public schooling and proposed a bill to create free schools in the state of Virginia. In support of this legislation, he reminded his fellow Virginians that anyone should have the opportunity to be elected to office, regardless of "wealth, birth, or

other accidental condition" and that a public educational system would ensure the best possible representatives in government (Kaestle, 1983, p. 6). New York Governor George Clinton echoed this sentiment, in 1782, when he implored his state's general education board, known as the Regents, to provide state funding for common public schools in order to avoid an existential crisis for the young country (Kaestle, 1983). Similarly, essayist Samuel Harrison Smith wrote in 1797 that "an enlightened nation," best maintains people's rights (Kaestle, 1983, p. 7).

These early ideas laid the groundwork for what came to be known as the common school movement. The movement gained traction in the New England region approximately fifty years after the Declaration of Independence, and aimed to provide a free education to all children - poor and rich, alike - so that all inhabitants could improve their lot in life (Reese, 2005). Like their Revolution-era predecessors, educational leaders of the 19th century believed that universal public schools were essential to democracy and that a public educational institution would help reduce social class divisions and tensions that had become prevalent in America (Reese, 2005). Horace Mann, one of the most prominent of these reformers, referred to universal public education as the "great equalizer," since he believed that schools could take children from diverse backgrounds, provide them with a common education and thus produce a balanced and unified society (Fuhrman & Lazerson, 2005). Anyone "stamped with inferiority" could be saved through education and rise to the "common level," Mann declared from the Massachusetts state senate floor during his tenure as a state senator (Goldstein, 2014, p. 24).

Of course, despite this egalitarian rhetoric, black slaves, their descendants and other marginalized groups of people, such as women and Native Americans, were excluded from this educational vision (Adams, 1995). Indeed, official desegregation of schools occurred well over a century after the common school movement's formation. These inequalities overshadowed the ideals of equality that were preached by early American educational leaders (Hochschild & Scovronick, 2003). Disparities in the educational and life outcomes of different racial, ethnic and socioeconomic groups that exist today reflect the inequities that have been present in the American public school system since its formation.

Recent data show that whites outperform or out-achieve blacks and other minority groups across nearly all aspects of American life, including, but not limited to: political participation, median family income, level of educational attainment and home ownership (Waters & Eschbach, 1995). Similarly, racial disparities in income attainment have grown steadily over the past fifty years, with black adults earning significantly less than their white counterparts (Putnam, 2015; Rothstein, 2017; Smeeding, 2005). These inequalities contribute to what Robert Putnam (2015) terms "de facto segregation" of Americans across class and racial lines (p. 27). Richard Rothstein, in his exploration of housing policies throughout America's modern history goes a step further and emphasizes the *de jure* segregation that black Americans continue to face as a result of the inequitably and systemic nature of discriminatory laws and public policy (Rothstein, 2017).

Student achievement data also highlight the separate and unequal outcomes for different population subgroups. According to the National Assessment of Educational

Progress (NAEP), by the time black students graduate high school, they are typically four years behind white and Asian students, meaning that a typical 12th-grade black student performs at the same level as a typical white student in eighth grade (Thernstrom & Thernstrom, 2003). In addition, minority students and inner city students have a higher dropout rate and lower rates of literacy and mathematical proficiency than white and suburban school students (Hochschild & Scovronick, 2003).

This gap, between white and black student achievement, has remained steady over the past 25 years (Hochschild & Scovronick, 2003, Thernstrom & Thernstrom, 2003). For example, in 1998, over 40 percent of black students scored below basic on the NAEP reading assessment, compared with just over 15 percent of white students. In 2000, nearly 70 percent of black students scored below basic on the NAEP math assessment, compared with just over 20 percent of white students. In 2015, the most recent NAEP administration, eighth-grade white students outperformed black students by approximately 32 points in math assessment and by over 25 points in reading, numbers that have remained nearly unchanged since the early 1990s. (https://www.nationsreportcard.gov/). As these examples illustrate, the equal and balanced society that Mann and other educational reformers hoped would follow from universal access to public education has clearly not yet come to pass. Instead, the opposite has occurred. Indeed, the American public school system has come to reflect and promulgate the inequalities that remain writ large in American society (Hochschild & Scovronick, 2003; Rothstein, 2017).

4

Statement of Problem

The roots of the persistent racial achievement gap are widespread and are based in the foundations of American society. Most importantly, the legacy of slavery, discrimination and racial oppression against blacks and other minority groups continues to have a negative impact on long term outcomes for black students today (Rothstein, 2017; Thernstrom & Thernstrom, 2003). Since this country's founding, blacks have been denied equal access to high quality education, job opportunities and the ability to live in more affluent neighborhoods. Discrimination and racial oppression persisted long after the abolition of slavery in the 1860s and past the Civil Rights movements of the 1960s and continues to reinforce the inequalities that have long been visible in American society (Hartman, 2015; Putnam, 2015; Rothstein, 2017; Thernstrom & Thernstrom, 2003).

Evidence of this historic legacy can be seen in many facets of American society. For example, black students and families are much more likely to live and go to school in high-density, high-poverty school districts and neighborhoods (Moore, 2004; Putnam, 2015; Rothstein, 2017). Black home environments are also likely to contain fewer books and other educationally enriching materials than white and more affluent family homes (Ogbu, 2004; Roscigno, 1998; Rothstein, 2017; Thernstrom & Thernstrom, 2003). Black and lower socio-economic parenting practices and cultural norms prioritize safety and compliance over intellectual curiosity and risk-taking (Putnam, 2015; Rothstein, 2017). Enrollment of black students in advanced academic courses is substantially lower than white student enrollment at all educational levels (Putnam, 2015; Smith et al., 2017; Thernstrom & Thernstrom, 2003). By contrast, black students, and particularly black

male students, are more likely to be disproportionately labeled with disabilities requiring special education services and continue to face the brunt of harsh and exclusionary discipline practices in schools (Grant, 2014).

The practice of tracking or ability grouping in schools also makes disparities between black and white students evident, since blacks and other disadvantaged minority student populations, are disproportionately placed in lower-ability tracks when compared with their equivalent white peers (Kao & Thompson, 2003; Slavin, 1990; Southworth & Mickelson, 2007). In 1997, the Digest of Education Statistics, in a nationwide survey, stated that 46 percent of high school seniors who reported being in an advanced or college prep track were white, compared to 36 percent of blacks and 31 percent of Hispanics. By contrast, 56 percent of students who reported being in a general track were Hispanics, 49 percent were blacks and 43 percent were whites (Kao & Thompson, 2003). This trend has shown little signs of ebbing. For example, in 2007, in Stanford, Connecticut public schools, only 5.5 percent of black students were enrolled in middle school honors classes, compared to 78.7 percent of white students enrolled in honors classes (Burris, 2014). This imbalance is compounded by the fact that predominantly white and more affluent schools generally offer more higher level classes and have a larger percentage of students taking advanced classes than low-income, predominantly minority schools (Kao & Thompson, 2003). As a result of these discrepancies, there exists widespread agreement among educational experts that tracking plays a role in reinforcing racial inequality in America today (Gamoran & Mare, 1989; Roscigno, 1998; Southworth & Mickelson, 2007).

Purpose of the Study

The purpose of this study is to explore whether and how tracking structures in mathematics courses at the middle school level relate to differences in achievement between white and black students. Given the history of the American educational system, the persistent nature of the achievement gap and challenges raised by tracking structures, this study is particularly significant for educational reform. In the United States, most students are tracked into different math levels beginning in middle school or earlier (Cleary & Chen, 2009; Hanushek & Wossman, 2005). Middle school math often then serves as a "gateway" class that has major implications for a student's future learning trajectory and college and career pathways (Akos, Shoffner, & Ellis, 2007). Research shows that students who take advanced math courses in grades six through eight, typically culminating in algebra in eighth grade, are more likely to take and pass more advanced math classes in high school, are more likely to enroll and succeed in a four-year college and are more likely to pursue careers in a science, technology, engineering or mathematics (STEM) field (Adelman, 1999; Burris et al., 2006; Lee et al., 2010)

Prior studies have been conducted to determine the impact of tracking in mathematics on academic outcomes (Akos et al., 2007; Burris et al., 2006; Mason et al., 1992). This research has shown that students who are placed in advanced classes, or in heterogeneous classes of mixed ability, generally perform better, or no worse than, students who are tracked into lower level or homogenous math groupings (Burris et al., 2006; Mason et al., 1992). Though some studies indicate a slight increase in achievement outcomes for higher ability students in tracked settings, this rule is generally true for

students at all ability levels (Kao & Thompson, 2003; Slavin, 1990). For example, students identified as average, who were placed in advanced classes were ultimately more likely to enroll in more challenging and advanced classes later in their academic careers than average peers who were placed in math courses that supposedly better matched their average ability level (Mason et al., 1992). Students in advanced math classes also typically performed better on standardized assessments, even when controlling for prior ability levels and selection bias, than students who were enrolled in grade-level classes (Burris et al., 2006; Kao & Thompson, 2003; Leow, Marcus, Zanutto & Boruch, 2004; Mason et al., 1992; Slavin, 1990).

Starting in the late 1990s, several public school districts began experimenting with universal acceleration policies, which essentially mandated that all students enroll in Algebra courses by the ninth grade (Burris et al., 2006; Dougherty, Goodman, Hill, Litke & Page, 2015; Clotfelter, Ladd & Vigdor, 2015; Senechal, 2014). The department of education in the state of California, for example, incentivized local school districts to adopt acceleration policies. Similarly, the city of Chicago mandated acceleration policies for math and English across the school district. Though the level of enforcement of these acceleration policies has varied across districts and student achievement data have thus far been mixed, enrollment in advanced math courses has steadily increased across the country (Loveless, 2009; Senechal, 2014).

Research Questions

This study will examine the impact of tracking students in mathematics on middle school student achievement outcomes by comparing the achievement outcomes of students enrolled in advanced mathematics classes, with students of comparable ability

who are enrolled in grade-level math classes. Achievement outcomes for this study will be measured using benchmark data from the Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) assessment in mathematics. Research questions for this study include the following:

1. How do the achievement outcomes of students enrolled in advanced math classes compare to the achievement outcomes of students with comparable ability levels and background characteristics who are enrolled in grade-level math classes?

   a. Over the course of one academic year, based on MAP assessment data, do students in advanced math classes achieve more or less than statistically similar students who are enrolled in grade-level courses?

2. How do the achievement outcomes of black students who are enrolled in advanced math classes compare to the achievement outcomes of black students with comparable ability levels and background characteristics who are enrolled in grade-level classes?

   a. Over the course of one academic year, based on MAP assessment data, do black students in advanced math classes achieve more or less than statistically similar black students who are enrolled in grade-level courses?

3. How do the achievement outcomes of white students who are enrolled in advanced math classes compare to the achievement outcomes of white students with comparable ability levels and background characteristics who are enrolled in grade-level classes?

a. Over the course of one academic year, based on MAP assessment data, do white students in advanced math classes achieve more or less than statistically similar white students who are enrolled in grade-level courses?

Significance of the Study

Two questions underpin much of the contemporary research on math tracking. First, researchers seek to understand whether tracking students is an effective strategy to maximize student achievement outcomes in comparison to heterogeneous, or detracked, grouping. Typically, when investigating this problem, researchers look at achievement data for comparable samples of students to determine whether students who are tracked perform better over time than students who are not. Studies break students down into subgroups such as low, average and high achievers, and use statistical analyses to compare the benefits and disadvantages tracking provides for each (Akos et al., 2007; Burris et al., 2006; Gamoran & Mare, 1989; Leow et al., 2004; Loveless, 2009; Mason et al., 1992; Slavin, 1990). Studies that attempt to answer this question also focus on whether detracking students by universally accelerating them, leads to higher achievement for all students. In these studies, researchers analyze achievement data from before and after the implementation of universal acceleration to determine if there are any statistically significant differences (Burris et al., 2006; Hanushek & Wossman, 2006; Leow et al., 2004).

The second primary focus of math tracking research is correlation and causality. Specifically, researchers have attempted to identify the factors that contribute to and predict the improved or decreased student achievement of students in tracking versus

heterogeneous group settings. Researchers use both quantitative and qualitative approaches to look at a broad array of factors that correlate to student achievement, including teacher expectations, national educational policies, the social context of the classroom, parent and peer support and the impact that stereotype threat may have on student performance (Hanushek & Wossman, 2006; Mason et al., 1992; Rice, Barth, Guadagno, Smith & McCallum, 2013). Understanding these factors is critical to being able to apply successful practices more broadly across school systems and student populations.

A growing body of empirical research related to tracking has yielded an emerging consensus around the impact of tracking on student achievement, and has identified factors that lead to greater student achievement in accelerated math classes. Nevertheless, additional research is warranted to fully explore and understand nuances within the prior literature. For example, scholars have difficulty discerning between the impact of tracking on student achievement as compared to the impact of other variables, such as socioeconomic status and prior ability (Hanushek & Wossman, 2005; Loveless, 2009; Mason et al., 1992). In addition, study findings lack conclusiveness in terms of how tracking structures impact the academic achievement outcomes of students of different ability levels. Finally, the prior literature would be strengthened by additional studies that continue to identify some of the causes and correlations that exist between student achievement and academic outcomes.

This study aims to add to the literature on tracking and the achievement gap by providing additional data on whether and how tracking structures affect student achievement outcomes. Prior studies have determined that the impact of tracking on

achievement outcomes for higher ability students can be slightly different than the impact of tracking on students of average or lower ability students (Kao & Thompson, 2003; Mason et al., 1992). This study will expand on this previous work, and will examine how race, in addition to student ability levels, may also correlate to differences in student achievement outcomes when students are tracked by ability. The study design, as well as the study setting, also ensures a unique data set from which to assess the impact of tracking on student achievement. Finally, the statistical analysis methodology that will be used in this study will help to determine the causal and correlational nature of the relationship between the variables explored in this study.

Many educational leaders and policy makers consider the racial gap in achievement to be "the most important civil rights issue of our time" (Thernstrom and Thernstrom, 2003, p. 274). Tracking practices also align closely with questions of equity, opportunity and fairness, issues which are at the heart of the American educational system and democracy. Ultimately, therefore, this study will provide educational leaders, policymakers and school community members with guidance as they work to confront the achievement gap and determine the future of tracking structures and detracking initiatives at local, state and federal levels (Burris, 2014; Hochschild & Scovronick, 2003; Ogbu, 2003; Reese, 2005; Senechal, 2014).

Organization of the Dissertation

Chapter I describes the problem to be explored through this research as well as its importance. Chapter I also defines the study's research questions, limitations, and a definition of terms used throughout the research.

Chapter II provides a literature review on topics relevant to this study. The review begins with a historical summary of the achievement gap in the United States and how the context of the achievement gap has changed over time. The review then continues with a deeper exploration of tracking practices in the United States and examines research that has been conducted around tracking in general and in mathematics education specifically. This part of the literature review also includes a review of pertinent studies conducted in support of and against detracking practices. The literature review then concludes by summarizing relevant prior research conducted in the Shaker Heights City School District, since this research provides a useful context to understand this current study.

Chapter III includes the study's methodology, data collection procedures and variables. Chapter IV includes research findings from the study focused on each research question. Chapter V summarizes the study and its results, concluding with a discussion of results and their implications for researchers and practitioners.

Limitations of the Study

The following limitations must be considered as relevant factors when interpreting results:

1. Since 2014, the author of this study has served as an administrator in the district in which this study takes place. Care will be taken to identify and address any influence this subjectivity may have on methodology, yet the experience and background knowledge of the author will be utilized to help shape the understanding and discussion of results.

2. While the sample size that will be used in the study is large enough to achieve statistically significant results and to extrapolate from these results conclusions that will aid researchers and practitioners in the field, it is important to note that all of the data that will be collected in this study will come from one school building in one inner ring suburb outside of a Midwestern city. Readers of this study should take this context into account and are encouraged to review additional studies that collect data from other similar and dissimilar school districts.

3. It is possible that additional factors, that will not be reviewed in this study, and which data will not be collected around, such as level of teacher experience, student mobility and parent levels of education, could also relate to and have an impact on student achievement outcomes in different course levels.

4. This study will attempt to address the issue of selection bias through its statistical methodology. Nevertheless, in this type of quasi-experimental design, selection bias may be related to differences in student achievement outcomes (Leow et al., 2004).

Definition of Terms

Achievement gap. The achievement gap most commonly refers to differences in academic performance between various student demographic groups. Typically, academic performance for purposes of the achievement gap is measured on high-stakes exams, such as state assessments in reading, mathematics and other content areas, the NAEP assessment and other norm-referenced assessments such as the NWEA MAP

measure (Anderson, Medrich & Fowler, 2007). Academic performance can also be measured in terms of student grades, grade point averages and other measures of learning. Historically, and for purposes of this study, the achievement gap most often refers to differences in achievement scores between white and black students (Anderson et al., 2007).

Black-white test score gap. The black-white test score gap refers specifically to the fact that black students score lower than white students on high-stakes national, state and norm-referenced tests in reading, mathematics and vocabulary. Black students also score lower than white students on measures of scholastic aptitude and intelligence (Jencks and Phillips, 1998).

Opportunity Gap. The term opportunity gap refers to the disparate opportunities that different student demographic groups have access to in the public education system. For example, black and other minority students are less likely to have access to higher quality teachers, more rigorous expectations in the classroom and better-resourced schools (Flores, 2007). This term is often used as an alternative to frame the discrepancies that exist in the achievement gap.

Income achievement gap. The income achievement gap describes the gaps in student achievement that exist between students at different socioeconomic levels. One measure of the income achievement gap is the average achievement difference between a child from a family at the 90th percentile of the family income distribution and a child from a family at the 10th percentile, also known as the 90/10 income achievement gap (Reardon, 2011).

Advanced course gap. The advanced course gap refers to the ratio of white students enrolled in advanced mathematics courses compared with black students enrolled in advanced mathematics courses. Lee (2002) calculated this ratio for 17-year old students who reported having taken algebra I, geometry, algebra II, pre-calculus or calculus according to the NAEP student survey.

Tracking. Tracking refers to the practice of placing students in different course levels, based on perceived abilities, background experiences and potential career paths. In today's schools, tracking most often takes place when students are placed, or enrolled in, different course levels, such as advanced, or honors level English, as compared with grade-level English. Tracking can begin as early as elementary school and typically continues through high school (Gamoran & Mare, 1989; Slavin, 1990).

Leveling. Leveling is typically used as a synonym for tracking (Burris, 2014).

Ability Grouping. Ability grouping is a close cousin of tracking and can happen between or within classrooms. Ability grouping occurs when students are placed in inter- or intra-class groups based on perceived abilities, background experiences or potential career pathways. At the elementary school level, ability grouping most often starts within a classroom. As students move up to higher grade levels, ability grouping can take place both within and across classrooms, as students are tracked into different course levels. At the high school level, ability grouping may also occur as students select specific courses in order to pursue specific career trajectories (Hoffer, 1992; Oakes, 1992).

Middle School. Sometimes referred to as middle level education, or junior high school, middle school typically refers to school for young adolescents. In the United

States, middle school usually includes some combination of grades six through nine (National Middle School Association, 2003).

CHAPTER II

LITERATURE REVIEW

This chapter consists of a literature review that focuses on three key aspects of this study: 1) the black-white achievement gap, 2) school tracking structures, particularly in the area of mathematics and, 3) research that has been conducted in and about the Shaker Heights City School District, the setting for this study. Prior research shows that tracking structures contribute to inequities in student achievement. At the same time, a close examination of the literature shows that additional research is warranted to better understand the intersections between race, student ability level and tracking and how these factors contribute to differences in student achievement outcomes.

The Black-White Achievement Gap and Its Significance

The 1954 Brown v. Board of Education Supreme Court decision to desegregate schools marked a new era in the American public education system, and gave hope to educational leaders and social scientists that racial disparities in academic performance would be a thing of the past (Slavin & Madden, 2006). Yet the publication of James Coleman's seminal report, *Equality of Educational Opportunity*, a mere twelve years later in 1966, disabused educational leaders of these aspirations. The Coleman Report highlighted persistent inequities in student outcomes across racial groups, a phenomenon which entered the popular American educational lexicon and became known as the

achievement gap (Clotfelter, Ladd & Vigdor, 2009; Fryer & Levitt, 2004; Lee, 2002). Since that time, educational leaders, researchers and policy makers have dedicated extensive time and energy to discover the underlying causes of the achievement gap, as well as to better understand the factors that contribute to the gap and how these discrepancies might be addressed and rectified.

In contemporary educational research, the achievement gap refers most often to differences in scores on state or national achievement tests between various student demographic groups. While discrepancies in academic achievement exist between nearly all racial subgroups, much of the research and attention to the achievement gap problem has focused on the gap in academic performance outcomes between black students and white students (Anderson, Medrich & Fowler, 2007; Clotfelter et al., 2009). For purposes of this study, this gap will be referred to interchangeably as both the black-white achievement gap and the black-white test score gap.

Until the 1960s, studies of the test-score gap between black and white students largely depended on samples of convenience. This practice changed in 1965, with the Equality of Educational Opportunity (EEO:65) survey. The EEO:65 survey was the first large scale statistical study that measured student performance across the nation and included a measure of race (Jencks & Phillips, 1998). Other similarly large surveys from around the same time period, include the National Longitudinal Study of the High School Class of 1972 (NLS: 72), the High School and Beyond Survey of 10th- and 12th-grade students conducted in 1980 and 1982, the National Longitudinal Surveys of Youth of 1980 and the National Education Longitudinal Survey of 1988. All of these surveys showed that black students scored lower than white students on a range of state and

national assessments in vocabulary, reading and mathematics, as well as on tests that claim to measure aptitude and intelligence (Jencks & Phillips, 1998).

The passage of the Elementary and Secondary Education Act (ESEA) of 1965 led to the formation, in 1969, of the National Assessment of Educational Progress (NAEP) to monitor and measure the academic achievement of students as they progressed through the American public education system (Vinovskis, 2001). Commonly referred to as "The Nation's Report Card," NAEP data have become the most often cited national measure of the achievement gap (Barton & Coley, 2010; Lee, 2002). Over the past three decades, researchers have been able to use nationally representative data, such as NAEP, to measure the extent of this achievement gap in more depth. Indeed, research on student achievement outcomes from the past fifty years demonstrate in no uncertain terms that the black-white test score gap is a "robust empirical regularity" (Fryer & Levitt, 2004, p. 447). While not the focus of this study, the black-white achievement gap also exists beyond test scores, in terms of dropout rates, numbers of students taking advanced courses, and in college admission rates (Ladson-Billings., 2006).

On the NAEP assessment, the typical black student scores below 75 percent to 85 percent of white students (Jencks & Phillips, 1998). Put another way, black students in the 12th grade generally perform as well as white students in the eighth grade. In addition, by the time they graduate high school, 91 percent of black students and 87 percent of Latino students are deemed not proficient in mathematics. By contrast, only 63 percent of white high school seniors and 53 percent of Asian high school seniors are deemed not proficient (Flores, 2007). In 1998, 43 percent of black students fell below the basic level of proficiency in reading, as compared to only 17 percent of white students

(Thomas & Brady, 2005). Given these numbers, it is not surprising that scholars and politicians frequently refer to the black-white achievement gap as "the most important of all educational problems" in the United States (Slavin & Madden, 2006, p. 389).

Metrics that describe the achievement gap typically show the number of standard deviation units' difference that exist between black test-score performance and white test-score performance. Starting around 1973, the gap between 17-year-old black-student performance on the NAEP assessment in mathematics was slightly more than one standard deviation lower than white-student performance. This gap narrowed by 20 percent to 40 percent over the course of the next 15 years, reaching its narrowest point in approximately 1989, when the gap was closer to .6 standard deviation units (Barton & Coley, 2010; Jencks & Phillips, 1998; Lee, 2002). From 1989 to the present, however, the gap has widened back to approximately one standard deviation difference (Fryer & Levitt, 2004; Harris & Herrington, 2006). Similar trends exist in NAEP reading scores (Barton & Coley, 2010; Jencks & Phillips, 1998).

In terms of raw scores, in 1973, 13-year old black students scored 46 points lower, on average, than their white counterparts on the NAEP assessment in mathematics. This number narrowed to around a 27-point differential in the mid- to late-1980s and widened back to a 32-point spread in 1999. The gap then narrowed slightly, to approximately 28 points in 2008. On the 1971 NAEP reading assessment, 13-year old black students scored 39 points lower than white students, on average. By 1988, this number dropped to an all-time low of 18 points and then widened again to 32 points in 1996. In 2008, the difference in black and white scores on the reading assessment was measured at 21 points (Barton & Coley, 2010). Most recently, on the 2015 NAEP

administration, black high school seniors scored 30 points lower, on average, than white students in both mathematics and reading. For eighth-grade students, this gap was 32 points in mathematics and 26 points in reading (https://www.nationsreportcard.gov).

Research on the achievement gap issue has attempted to determine whether discrepancies in academic performance between black and whites narrow, persist or expand as children progress through school. Answering this question is important, because it may help shed some light on the role and power of schools to change the achievement gap trajectory. Scholars generally agree that the achievement gap exists before children start kindergarten and that it continues through adulthood (Fryer & Levitt, 2006; Jencks & Phillips, 1998). From there, however, opinions diverge as to how much the gap widens over a child's educational career. At best, the test score gap remains roughly constant from kindergarten through 12th grade (Clotfelter et al., 2009). More likely, however, the achievement gap widens, with some scholars pointing to evidence that black students enter elementary school one year behind white students, but lag behind three to four years by 12th grade (Jencks & Phillips, 1998; Fryer & Levitt, 2004).

Longitudinal study data and an in-depth analysis of student achievement data paint a more nuanced portrait of the achievement gap trajectory over the course of a child's schooling. Some data show that math and vocabulary gaps widen between first and 12th grade, whereas the reading gap remains fairly consistent over time. Other studies indicate that a black student who starts off elementary school with approximately the same test score performance as his or her white counterpart typically finishes elementary school with similar math scores, but with lower scores on reading and vocabulary assessments. Similar studies indicate that a black student who starts high

school with the same test score as his or her white counterpart typically finishes high school with the same even score (Jencks & Phillips, 1998). Finally, white students who start elementary school with tests scores at the mean, are likely to complete high school with test scores at the mean. Black students, however, who begin elementary school with test scores at the mean are likely to decline to .35 to .40 standard deviations below the mean by the time they are high school seniors. This decline would represent a slide of 35 to 40 SAT points over a student's academic career (Jencks & Phillips, 1998).

Studies that attempt to measure the growth of the black-white test gap as children progress from kindergarten through high school typically control for a wide range of covariates. For example, Fryer and Levitt (2004) conducted a study of raw test scores using data from the Early Childhood Longitudinal Study kindergarten cohort (ECLS-K). Covariates included factors such as the child's age, the mother's age, number of books in the home, level of parental educational attainment and socioeconomic status (Fryer & Levitt, 2004). Studies that control for these covariates show that when comparing similar black and white students, black students score only slightly worse in math when entering kindergarten than their white peers. Nevertheless, by the end of the third grade, the black-white test gap "is evident in every skill tested in reading and math," even when controlling for these covariates and others (Fryer & Levitt, 2006, p. 252).

A strong correlation exists between the black-white student achievement gap and disparities between black and white outcomes later in life (Lee, 2002). For example, studies show that gaps in wages between white and black adults parallel the same gaps that are seen in eighth-grade test scores (Fryer & Levitt, 2004). Data from the National Longitudinal Survey of Youth (NLSY) also show that black male workers who test at or

above the 50th percentile on the Armed Services Vocational Aptitude Battery (ASVAB),

earn wages that are nearly equivalent to the earnings of the average white male worker.

By contrast, the wages of black male workers whose test scores fell below the 50th

percentile are significantly lower than the average white male worker (Jencks & Phillips,

1998).

Gaps in achievement also mirror inequalities in employment rates. In 2010, the

employment rate for black males with 12 years of school was 68 percent, compared to 87

percent for white males with the same level of schooling. Yet black adult males with

higher levels of education maintained an employment rate of 89 percent as compared to

95 percent for whites (Barton & Coley, 2010). These data show that the achievement gap

holds real implications for the ability of black males to find jobs at a rate that is

consistent with white males of similar achievement levels.

Factors Contributing to the Achievement Gap

Scholars have conducted an extensive amount of research to determine the factors

that contribute to the black-white achievement gap. Prior literature focuses on factors

such as environmental influences, school structures and composition and family

background. This research has led to a greater understanding of the causes and ongoing

nature of the achievement gap. Nevertheless, despite the wide-ranging nature of

achievement gap research, gaps remain in the general understanding of why the

achievement gap exists and how to address it.

Any notion that discrepancies in student achievement between white and black

students is tied to immutable genetic differences between racial subgroups has been

broadly disproven (Barton & Coley, 2010; Fryer & Levitt, 2006; Gamoran, 2001; Jencks

24

& Phillips, 1998). For example, research on intelligence and aptitude testing shows that environmental context has a statistically significant impact on score results. In addition, studies that have examined the achievement outcomes of black children raised in white homes show general increases in test scores, which support the argument that environmental factors may play a role in achievement outcome metrics. Finally, trends in the achievement gap, such as the decrease in the gap that took place during the second half of the twentieth century, indicate that larger societal forces play a role in the contours of the achievement gap. "The black-white test score gap," therefore, "does not appear to be an inevitable fact of nature" (Jencks & Phillips, 1998, p. 2). Nevertheless, there are long standing historical, political and cultural factors that drive the persistent nature of the achievement gap.

Disparities in the educational achievement of black students as compared to white students stretches back to the roots of the black experience in the United States. Thernstrom and Thernstrom (2003) call this phenomenon the "cultural inheritance" of blacks, noting that the modern black life in America is shaped by "a very long history of racial oppression – centuries of slavery, followed by disenfranchisement, legally mandated segregation, and subordination in the Jim Crow South and intense racial prejudice in the North" (p. 121). Ladson-Billings (2006) frames this legacy of institutionalized slavery, racism and prejudice as a "historic debt," and asserts that black academic achievement remains hampered by past inequities (p. 5). For example, African-Americans were forbidden to be educated during the period of slavery and black students in the south did not have access to universal public secondary school education until 1968 (Ladson-Billings, 2006). In addition, redlining practices in the 1960s

prevented black families from moving to more affluent suburbs and created a segregated system of residential areas, that is still evident today (Barton & Coley, 2010).

This historic legacy means that black family median income remains lower than white families and that black families typically reside in higher poverty areas than their white counterparts. Barton and Coley (2010) consider these circumstances as analogous to being hit with "a triple whammy in the home, neighborhood and school" (p. 33). According to the authors, black children, on average, are "impaired in their development, lack family capital, and face hostile neighborhood environments. They are also likely to attend lower-quality schools staffed by lower-quality teachers" (Barton & Coley, 2010, p. 33). Partly as a result of these external factors, black student achievement remains persistently lower than white student performance.

The 1965 Moynihan Report predicted that the black legacy of slavery and discrimination would have distressing implications for the health of the African-American nuclear family. In particular, warned Moynihan, "the single-parent rate" of black families, "would continue to rise unless the nation did something positive" (Barton & Coley, 2010, p. 21). According to Moynihan's projections, black families, and particularly black males, would bear the brunt of high unemployment rates, ongoing discrimination and a poorly structured American welfare system (Moynihan, 1965).

Unfortunately, Moynihan's predictions for black families have largely come to pass. The steep rise in black children being raised without fathers coincides with the lack of progress made in narrowing the achievement gap (Barton & Coley, 2010, p. 24). There are also significant gaps in employment rates, particularly for black males. In 2010, the employment rate for black males with 12 years of school was 68 percent,

compared to 87 percent for white males. By contrast, the reverse is true in terms of the number of adult males who have been incarcerated. Nine percent of black males between the ages of 26 and 30, with nine to 11 years of education and who were born between the years 1950 and 1954, were incarcerated. This number goes up to 19 percent for black males of the same age, with the same level of education, who were born between 1960 and 1964 and increases again to 26 percent for the same demographic subgroup born between 1970 and 1974. Astoundingly, for white males of the same corresponding demographic subgroup, the percentages are two, four and five respectively (Barton & Coley, 2010, p. 25)

The history of the black experience in America has also had an impact on culture and home life. Research has shown that black families have less books at home than white families and that black students are more likely to watch more television after school hours than white students (Ladson-Billings, 2006; Ogbu, 2003). In addition, black students are more likely to feel isolated in the classroom and are less likely to engage in collaborative study practices that are shown to have benefits on achievement outcomes (Steele, 2011). Studies such as Ogbu's examination of black students in an affluent suburb also point to the notion that black students may have lower academic performance because being academically successful is seen as acting white (Gamoran, 2001; Ogbu, 2003). Finally, black students may be subject to harsher, more hierarchical discipline at home that focuses on being safe and complying with rules, rather than self-advocacy and negotiation (Putnam, 2015).

Beyond these deeply entrenched historical and cultural factors, the scope and changing contours of the black-white achievement gap over time provides some

additional clues about discrepancies in educational outcomes. From the 1970s until the mid-1980s, NAEP data show substantial improvements in black and Hispanic achievement outcomes and a corresponding narrowing of the achievement gap. This narrowing of the achievement gap in the 1970s and 1980s primarily occurred because of increases in black achievement levels, while white student achievement level remained flat. When the gap started growing in the late 1980s and through the 1990s, white student achievement rose, while black student achievement stayed flat. SAT score gaps during this time period show similar trends (Lee, 2002).

The narrowing of the achievement gap in the 1970s occurred simultaneously with the rollout of major policy changes and social initiatives across America. By the 1970s, the *Brown v. Board of Ed.* decision had forced school systems to desegregate, and the quality of education afforded to minority students had therefore improved, as compared to the prior system of segregated schooling (Harris & Herrington, 2006). One sign of the improvement in black student access to high quality education was that, beginning in the 1970s, there was an upward trend in the number of black students enrolled in advanced courses (Barton & Coley, 2010; Lee, 2002). In addition, Lyndon Johnson's Great Society legislation provided material support to schools to attempt to better meet the needs of students from low socioeconomic backgrounds. For example, the Elementary and Secondary Education Act (ESEA) of 1965 supplied federal Title I funding to schools based on the percentage of low-income students who attend the school (Slavin & Madden, 2006; Thomas & Brady, 2005). Project Head Start was also initiated at this time and created early childhood interventions and education for lower income families (Barton & Coley, 2010; Jencks & Phillips, 1998).

During this time period, other quality of life indicators for black students and families also moved in the positive direction. In the 1970s and 1980s, more parents of color attained higher degrees and earned more income than in previous time periods. More black women also found employment, were married to the father of their children and began having children at older ages (Barton & Coley, 2010). Research shows that the positive movement in these environmental factors likely had an impact on the narrowing of the achievement gap in the 1970s and 1980s. When the achievement gap began growing again in the 1990s, however, black family conditions remained the same or continued to improve. As a result, researchers began focusing on other factors that contribute to the achievement gap (Lee, 2002).

In the late 1980s and 1990s, higher-performing white students experienced the biggest increase in academic achievement outcomes. These results correspond to a push for academic intensity and more rigorous learning standards that took place around the same time (Harris & Herrington, 2006; Lee, 2002). In 1983, the Reagan administration published *The Nation at Risk*, a report that warned of the dire consequences of a school system riddled with inequity and that was failing to prepare students for the modern workforce (Harris & Herrington, 2006; Thomas & Brady, 2005). Consequently, the federal government's attention turned to the raising of academic standards for all students and encouraged school districts across the country to develop more rigorous subject area standards and to ensure that curriculum and assessments aligned to these new standards (Carnoy & Loeb, 2002; Domina & Saldana, 2012). The federal government also amended Title I to connect the receipt of funding to student achievement results, and required states to annually assess student academic progress based on high stakes,

standardized test scores (Harris & Herrington, 2006; Thomas & Brady, 2005). At the same time, school districts began creating more uniform standards for teaching certification and raising the requirements to graduate high school (Domina & Saldana, 2012; Thomas & Brady, 2005).

Despite the coincidental nature of the academic intensification movement and the growth in the achievement gap, it remains unclear exactly why more rigorous standards led to growing disparities between black and white student outcomes. Nevertheless, scholars have found that academic differences become manifest as children get older and as students are required to engage with higher level questioning and learning tasks (Fryer & Levitt, 2006). Around the time of academic intensification, black student dropout rates also began increasing. By contrast, the dropout rate for white students - which has always been lower - began decreasing (Lee, 2002).

In many ways, the recent federal accountability movement mirrors the academic intensification period of the 1980s and 1990s. At the beginning of 2002, Congress passed the No Child Left Behind Act (NCLB), which required school districts to ensure that all children reach grade-level proficiency on high stakes assessments by the 2013-2014 school year (Kane & Staiger, 2002). In addition, in order to address the achievement gap issue, NCLB required schools and districts to demonstrate that all statistically significant subgroups within a school's demographic were making Adequate Yearly Progress (AYP) (Andersen et al., 2007; Thomas & Brady, 2005). Finally, federal incentive grants, such as Race to the Top funding, encouraged states to adopt new, more rigorous, Common Core subject standards, and to develop more thorough and demanding teacher evaluation systems (McGuinn, 2016). Data from the NCLB time period show that achievement

30

scores for both black and white subgroups have recently improved. The gap between these two subgroups, however, has remained largely the same (Barton & Coley, 2010).

In addition to academic intensification efforts, prior research on the achievement gap has focused on the changes in the racial composition of school demographics in the late 20th century. The desegregation trend that had taken off in the 1960s was reversed in the last decade of the 20th century. By 1997, for example, 69 percent of black students attended minority-majority schools, where minorities comprised the majority of the school population, as compared to 63 percent in 1987 (Lee, 2002). Scholars surmise that the re-segregation of schools amplified the effects of other factors that correlate to student achievement, such as teacher expectations, school resources and access to challenging curriculum (Jencks & Phillips, 1998; Moore, 2004). In addition, factors such as the percentage of students eligible for free or reduced meals, gang incidents, hall pass policies and Parent/Teacher Association funding indicate that majority-minority schools are generally of lower quality than schools with a higher percentage of white students. Thus, these "systematic differences" in school quality for blacks and whites likely account for some of the discrepancies in student achievement outcomes (Fryer & Levitt, 2004, p. 457).

In a similar vein, segregation within a school, under the guise of ability grouping or tracking, correlates with racial gaps in student achievement outcomes. Black students, for a variety of reasons, are not evenly represented in higher-level ability groups and courses. Yet students in higher-level groups and classes tend to outperform their peers. Jencks and Phillips (1998) report that "enriched and accelerated classes probably do increase the test score gap between high and low scorers, since they benefit students who

31

already score high" (p. 335). Students in higher-level classes and groups are exposed to more rigorous instruction and curriculum, while lower-level students are subject to lower teacher expectations and perceptions of their academic ability. Similarly, minority students are more likely to be placed with teachers who are less prepared, have less experience and who turn over more frequently than teachers of white students (Harris & Herrington, 2006). Research shows that as these experiences accumulate from kindergarten through high school, they have a substantial impact on student achievement outcomes (Gamoran, 2001; Jencks & Phillips, 1998).

Studies show that hundreds of different family characteristics also correlate with children's test scores. One series of studies, conducted using data from the National Longitudinal Survey of Youth (CNLSY) measured children's aptitude using the Peabody Picture Vocabulary Test (PPVT). The studies found that the level of parental schooling, and particularly a mother's level of educational attainment, has a statistically significant relationship with student test scores. The same studies also found that parenting strategies have a statistically significant relationship with test score performance (Fryer & Levitt, 2006). Strategies that are associated with middle class attitudes and behaviors, such as less punitive discipline in the home and limitations on screen time, are generally found to have a positive impact on student achievement outcomes (Lareau, 2011; Ogbu, 2003). Though an increasing number of black families are becoming middle class, black student achievement scores remain low, because generations of relatives who were raised outside of the middle class continue to influence family dynamics, thereby causing a lag in the adoption of middle class attitudes and practices (Jencks & Phillips, 1998).

32

As these studies illustrate, there is a strong correlation between socioeconomic differences and academic achievement disparities. More affluent children, who tend to be white, have parents with higher levels of education and more income, two factors that contribute to academic achievement. A measure of this correlation is the income achievement gap, which has been growing over the past fifty years (Reardon, 2011). The income achievement gap measures the average achievement difference between a child from a family at the 90th percentile of the family income distribution and a child from a family at the 10th percentile. This statistic is also known as the 90/10 income achievement gap. As a frame of reference, the income achievement gap in 2001 was approximately 30 to 40 percent wider than the income achievement gap 25 years earlier. In addition, the 90/10 income achievement gap is now nearly twice as large as the black-white achievement gap, whereas, the inverse was true fifty years ago. Like the black-white achievement gap, the income achievement gap is large when students enter kindergarten and remains relatively constant throughout a child's school progression (Reardon, 2011).

The income achievement gap is growing partly because more income leads to stronger student achievement, which leads back to more income. Higher income families invest more time and resources in their children's academic development than low income families. In addition, higher income families have access to more and higher quality socioeconomic and academic resources than lower income families, including higher quality schooling. Like racial segregation, studies show that increased income segregation, in both schooling and residential areas, contributes to these trends. As a

result, differences in family income now correspond to 30 to 60 percent more difference in academic achievement than in the 1970s (Gamoran, 2001; Reardon, 2011).

Despite the growing income achievement gap, studies that control for a wide range of covariates show some surprising results. For example, when researchers control for income, schooling and a mother's test scores, racial disparities in parental wealth were found to "have almost no effect on children's test scores" (Jencks & Phillips, 1998, p. 23). This finding means that income inequality alone does not explain the test score gap.

Similarly, having a single-parent household as compared with a two-parent household has no appreciable impact on test scores, when controlling for a mother's family background, test scores, and educational attainment. This finding indicates that a white student and a black student who both come from a single-parent household, and whose mothers come from the same type of family background, with the same level of schooling and with the same test scores, can expect to have similar achievement outcomes (Jencks & Phillips, 1998).

Some scholars believe that flaws in standardized tests may account for some of the discrepancies in achievement outcomes between black and white students (Lee, 2002). For example, blacks may do worse on exams because the test itself was not devised in a culturally proficient manner, or because black students have less developed test taking skills. This notion, that black students simply do poorer on high stakes exams than white students, corresponds with the idea that black students are generally not as successful at doing school as their white counterparts (Fryer & Levitt, 2004). Given the broad scope of the achievement gap, however, it seems unlikely that culturally deficient

tests, or differences in tests taking practices, can serve as the primary explanation for the black-white achievement gap.

As this review of prior literature demonstrates, factors that contribute to the black-white achievement gap include environmental influences, school demographics and socioeconomic status. This research has led to a greater understanding of the causes and ongoing nature of the achievement gap. Nevertheless, research has also shown that there is still work to do in order to acquire a more complete understanding of why the achievement gap exists and how to address it.

Closing the Black-White Achievement Gap

Scholars agree that addressing the achievement gap will require long-term, sustained efforts across many fronts (Barton & Coley, 2010). Beyond this basic consensus, however, scholars disagree about precisely what types of efforts will ultimately prove successful. Furthermore, like research on the achievement gap itself, gaps remain in a complete understanding of how different factors can help narrow or widen the achievement gap. Nevertheless, a review of the literature provides some recommendations for school reform in order to equalize achievement outcomes between black and white students.

Some scholars believe that only drastic change will lead to improved educational equity. These researchers assert that traditional ideas about how to close the black-white achievement gap, such as promoting more rigorous academic content standards and further desegregating schools, have not proven entirely successful and that other avenues must therefore be explored (Jencks & Phillips, 1998). Ladson-Billings (2006), for example, states that the only way to eliminate the achievement gap is to declare "moral

bankruptcy" and to then begin "from the ground up to build the kind of education system that would aggressively address" systemic inequalities (p. 10). In this way, new schooling options would be created that meet the needs of all students (Ladson-Billings, 2006).

Other researchers contend that educational leaders and policymakers must confront factors that cause black students to start kindergarten below their white peers (Jencks & Phillips, 1998). Barton and Coley (2010), for example, emphasize that improving the health of our "smallest schools," or individual families and neighborhoods, will lead to a closing of the achievement gap (p. 34). "The idea of a substitute for the institution of raising children is almost unthinkable," state the authors (Barton & Coley, 2010, p. 35). Other scholars focus on supporting black parents to raise their young children. Programs such as Head Start, for example, can provide black parents with guidance to support early childhood cognitive development in the home (Jencks & Phillips, 1998).

In the age of federal accountability measurers, researchers have also attempted to quantify whether government-driven school reform can be used to narrow the black-white achievement gap. This question is especially relevant in the face of recent major federal reform initiatives, such as the Every Student Succeeds Act, Race to the Top competitive grants and the No Child Left Behind Act. Scholars generally agree that government based accountability policies may increase overall student achievement, but it is not clear whether they help to narrow the achievement gap (Gamoran, 2001; Harris & Herrington, 2006).

Supporters of government accountability policies claim that government-based accountability helps put pressure on low-performing schools to offer a high quality education to minority and disadvantaged students. Additionally, public reporting mechanisms, such as state school report cards, identify weaknesses in schools which can then be addressed. Furthermore, public reporting generates political and economic pressure for schools to meet the needs of lower performing students in order to achieve or maintain high accountability grades (Harris & Herrington, 2006). Proponents of these measures also argue that the more stringent graduation requirements instigated by federal reform efforts may help reduce the achievement gap, since they provide incentives for all students to take challenging courses (Harris & Herrington, 2006). Critics of these measures, however, assert that accountability policies ultimately widen the gap, because high performing students tend to do better on the high stakes assessments that are key features of federal accountability programs (Gamoran, 2001; Harris & Herrington, 2006).

Prior research emphasizes actions that schools and school districts can take in order to narrow the achievement gap. Research shows that black students are less likely to have experienced, high quality teachers who challenge them to think critically about the content (Flores, 2007; Harris & Herrington, 2006). Black students are also less likely to have access to computers, or to use the computers in the classroom to complete higher level work. Finally, black students are more likely to be placed in lower-level classes than their white counterparts, even when the students have the same test scores (Flores, 2007). Given these statistics, some authors refer to the achievement gap as one of opportunity rather than of academic ability. These scholars contend that improving access to high-quality educational opportunities can narrow the achievement gap (Flores,

2007). In order to close this opportunity gap, all students should therefore have "equitable and optimal opportunities to learn...from a well-qualified teacher who will make connections to the background, needs and cultures of all learners" (Flores, 2007, p. 37).

Prior research also shows that placing students in racially integrated and socioeconomically diverse ability groups, classrooms and schools can narrow the achievement gap. As Coleman first explicated in his 1965 report, the socioeconomic composition of a school and classroom strongly correlates with student academic achievement. Coleman's data showed that the family background of a student's classmates had a strong correlation with student achievement (Ladson-Billings, 2006). "The social [class] composition of the student body is more highly related to achievement, independent of the student's own social background, then is any school factor (Coleman et al., 1966, p. 325 in Saporito & Sohoni, 2007, p. 1230). Though Coleman's report focused specifically on the socioeconomic status of a child's classroom peers, scholars agree that the racial diversity of a child's peers also plays a factor in student achievement (Saporito & Sohoni, 2007). Students are therefore most likely to do best when they are surrounded by racially and socioeconomically diverse peer and support networks (Ladson-Billings, 2006; Putnam, 2015).

An analysis of student performance in Raleigh, North Carolina supports the claim that heterogeneous grouping helps to address educational outcome disparities. Grant (2009) studied changes in student achievement that occurred after a relatively affluent and white suburban school district merged with Raleigh's poorer urban school district that had a high concentration of minority students. School leaders of the new

metropolitan district dedicated a great deal of effort to ensure that all new schools in the District housed a diverse blend of students. Student testing data from the newly formed district indicated that achievement outcomes improved significantly under the new district structure across most student subgroups (Grant, 2009).

Steele (2011) explored one reason that heterogeneous grouping leads to a reduction in the black-white achievement gap in his analysis of stereotype threat. Stereotype threat is the notion that students may perform badly on assessments when they self-identify with specific groups. For example, female students in an advanced math class may feel threatened by the common stereotype that girls are bad in math, and therefore end up doing worse than their male counterparts. According to Steele (2011), diverse classrooms are less subject to stereotype threat because students are less likely to be affected by a categorization with a specific racial or gender subgroup and because all students will find role models and high-performing peers with whom to associate.

Past research also indicates that both teacher quality and smaller class size correlate with improvements in black student achievement (Harris & Herrington, 2006; Jencks & Phillips, 1998). In fact, a teacher's own test scores appear to correlate strongly with student outcomes in his or her classroom. To narrow gaps in academic achievement, schools with a high concentration of minority students could factor information related to a teacher's test performance history into the hiring process, while also remaining committed to a diverse teaching corps. In addition, schools, particularly those with a high percentage of black children, should consider programming options that provide students with smaller classes (Jencks & Phillips, 1998).

Prior research also demonstrates that teachers generally have lower expectations for blacks than for whites, in terms of both academic and behavioral performance. Scholars of this phenomenon point to the self-perpetuating nature of these low expectations: because teachers have lower expectations for black students, they behave worse and perform worse than white students; teachers then base their lower expectations on past behavior and academic performance (Jencks & Phillips, 1998). In order to disrupt this cycle, school leaders should consider implementing professional development opportunities that provide teachers with opportunities to see disadvantaged black youth performing at high levels, thus reframing the perspective and potential biases that teachers hold of black students (Jencks & Phillips, 1998; Steele, 2011). In addition, school leaders should emphasize the importance of culturally relevant curriculum and instruction, in order to meet all students' needs (Jencks & Phillips, 1998).

Some studies focus on specific schools and programs that have achieved success in reducing the achievement gap. Balfanz and Byrnes (2006), for example, examined three high-poverty middle schools in Philadelphia. In their study, the authors found that schools that adopted comprehensive reforms to improve instruction and the school learning environment, combined with intensive teacher support saw gains in student math performance and a narrowing of the achievement gap for multiple cohorts of students (Balfanz & Byrnes, 2006).

Other researchers point to the success of some charter school networks, such as the Knowledge is Power Program and the New York-based Success Academy as evidence that improvements in minority student achievement can be accomplished through more time in school, elimination of barriers to reform, such as teacher unions,

and an intensive approach to curriculum and instruction that focuses on student test score results (Thernstrom & Thernstrom, 2003). The Success for All program, which is typically implemented in high-poverty, urban and minority school districts, has also achieved some measure of success by focusing intensively on reading skill development (Slavin & Madden, 2006).

Critics of these types of educational programs question whether these types of intensive supports are able to be widely replicated. In addition, researchers wonder whether these types of charter schools prioritize test score results over more authentic student learning (Sahm, 2015). Other researchers suggest that a closer examination of data is warranted in order to identify whether the gap is closing because of weaker performance by white students, rather than stronger performance by black students (Anderson et al., 2007).

Authors Jencks and Phillips (1998) state that "if racial equality is America's goal, reducing the black-white test score gap would probably do more to promote this goal than any other strategy that commands broad political support" (p. 4). Gamoran (2001) concurs and notes that unless the achievement gap is reduced, "persons in positions of power and advantage will use schooling to preserve their positions and those of their children" (p. 144). Prior research provides some pathways to address the achievement gap. Nevertheless, additional research is required to fully understand the potential impact of school- and system-wide reforms on the closing of the achievement gap.

Introduction to Tracking

Since the inception of the American public school system, educators have struggled with the question of how to best instruct and meet the needs of students who

enter a common school system with different levels of knowledge, abilities and background experiences (Southworth & Mickelson, 2007). In 1894, the National Education Association assembled the leading educators of the day, in a group that became known as the Committee of Ten, to chart a path forward on this issue and others confronting the growing public school system. "Should [a] subject be treated [taught] differently for pupils who are going to college, for those who are going to a scientific school, and for those who presumably are going to neither?" posed the committee (National Education Association of the United States, 1894, p. 17). In its final report, the members of the committee unanimously responded, "every subject which is taught at all in a secondary school should be taught in the same way and to the same extent to every pupil" (National Education Association of the United States, 1894, p. 17).

Though the Committee of Ten's report echoed the sentiments of equality expressed by educational and political leaders nearly a century earlier, the committee's reasons were actually based in the pragmatic realities of administering a growing national institution. If all subjects are taught similarly across schools, reasoned the committee, then secondary school curriculum across the country would be greatly simplified and there would need to exist a more uniform standard for teacher training, both of which were seen as crucial to the improvement of the American public school system (National Education Association of the United States, 1894; Southworth & Mickelson, 2007). Nevertheless, the committee's proposals faded as student enrollment in the American public school system expanded and as the student body became increasingly diverse.

In 1900, only six years after the formation of the Committee of Ten, the number of high school graduates as a percentage of the overall United States population doubled

to 6.4 percent. Twenty years later, this number rose to 16.8 percent, and by 1940, the number topped 50 percent (Bohan, 2003). Consequently, educational leaders began advocating for the division of students into different groups or academic tracks, based on perceived abilities, background experiences or career paths (Burris, 2014). Thus began the American public school system's history of tracking students into different course levels.

The confluence of several different factors led educational leaders to support tracking in public schools. Firstly, the growing industrial economy and corresponding assembly line mentality spread into schools, where educational leaders promoted a "cult of efficiency," to prepare different students for their different roles upon leaving school (Burris, 2014, p. 4). Leonard Ayres, an early 20th century educator, exemplified this approach in his aptly titled manuscript, *Laggards in our Schools*. Ayres declared that special programs were needed for students who fell behind grade level – which he claimed comprised the majority of the student population – so that schools could address the needs of the relatively small percentage of intelligent students for whom they were designed (Ayres, 1913).

Educational surveys conducted in the 1930s demonstrate the pervasiveness of these beliefs. The 1933 National Survey of Secondary Education, for example, noted that less than half of secondary schools required students to take algebra or geometry. A similar survey of teachers, taken a few years earlier, showed that over one-third of mathematics teachers felt that fewer students should take mathematics, rather than more (Stinson, 2004). These ideas and attitudes about education fell in line with the increasingly popular notion that schools needed to differentiate the curriculum for

different students in order to "prepare a diverse population for a range of societal needs" (Burris, 2014, p. 4).

Burgeoning immigration at the turn of the century also forced school leaders and administrators to figure out how to rapidly Americanize and provide English language instruction to non-native students. This influx of new and foreign pupils led many educators to conclude that separate tracks were necessary for different types of students (Burris, 2014). In recognition of these changes, the National Education Association published a new report in 1918, titled the "Cardinal Principles of Secondary Education." In contrast to the earlier Committee of Ten's report, this updated publication asserted that a democratic society is organized with different people taking on different roles and that public schools should be designed with this stratification in mind. "The school should develop the concept that the civic duties of men and women, while in part identical, are also in part supplementary," declared the 1918 report. As a result, "differentiation in civic activities is to be encouraged" (National Education Association of the United States, 1918).

During the first half of the 20th century, tracking at the secondary level primarily took the form of placing students in fully contained programs that predetermined all of the courses in which they were enrolled. Typically, this meant that upon entering high school, a student was placed in a college-preparatory track or a non-college preparatory track (Watanabe, 2007). Toward the second half of the twentieth century, however, tracking practices changed in response to evolving ideas and the forces of the standards movement. In what Samuel Roundfield Lucas termed the "unremarked revolution," beginning in the 1960s and 1970s, practice shifted so that students were placed into

different course levels, such as advanced placement, honors or regular courses, rather than in an overarching college or non-college prep program (Bernhardt, 2014; Southworth & Mickelson, 2007; Watanabe, 2007). This practice remains the norm today, with most students tracked into different level courses in mathematics, in English and in other core classes by the time they reach secondary school (Smith, Frey, Pumpian & Fisher, 2017).

Tracking is formally considered to include: placing elementary students in ability groups within and across classes; scheduling middle school students in classes according to their ability; and, establishing course trajectories in high school that prepare students for different post-secondary paths. All three of these practices are similar because they allow schools to teach similar students together, separate from other students, and because they allow teachers to use different teaching strategies and provide different learning experiences based on their deemed appropriateness for different levels of student (Oakes, 1992; Burris, 2014).

Ability grouping is predominantly used in math and science courses, but can also be applied to English and other subjects (Hochschild & Scovronick, 2003). Indeed, the history of tracking students specifically in mathematics can be traced back over 2500 years. In Plato's *Republic*, Greek philosophers Socrates and Glaucon engage in a discussion of mathematics and education. The two scholars agree that the study of mathematics should be reserved only for those who were "naturally skilled in calculation" (Sterling & Scott, 1996, p. 220). Furthermore, ordinary math students should study math in order to understand how to buy and sell goods, whereas those students who show innate ability and excel at math should "persist in their studies until they reach the level

of pure thought where they will be able to contemplate the very nature of number"

(Sterling & Scott, 1996, p. 219). In many ways, mathematics education in the American

public school system continues to reflect this stratified model.

Tracking in any school can be recognized in relation to four dimensions:

inclusiveness, selectivity, electivity and scope (Domina & Saldana, 2012). Inclusiveness

means the extent to which high level courses are available to students. Selectivity refers

to the extent that ability grouping creates homogenous learning environments. Electivity

relates to the degree to which students are able to choose their own course placements.

Finally, scope refers to the extent to which course placements in one subject area are

connected to course placements in another subject area (Domina & Saldana, 2012).

A range of criteria are typically used to place students in the appropriate ability

group or course. These criteria include both objective and subjective measures, which

some experts divide into "meritocratic" and non-meritocratic" categories (Bernhardt,

2014). Meritocratic criteria consist of objective parts of a student's academic record such

as standardized test scores, grades and other measures of prior achievement (Oakes,

2005). In theory, these criteria embody the notion that students who perform best based

on objective, unbiased data, should have access to the most advanced courses (Oakes,

1992). Some researchers, however, assert that these seemingly objective criteria do not

take into account group differences in initial entitlements between students (Roscigno &

Ainsworth-Darnell, 1999).

Non-meritocratic criteria are also used to make tracking placement decisions.

These types of factors include informal observations about a student's behavior and

motivation, teacher recommendations, parental preference, race and social class

(Bernhardt, 2014; Gamoran & Mare, 1989; Oakes, 2005). Though the specific criteria for placement and enrollment may vary across schools, scholars agree that "highly subjective" criteria frequently play a key role in these life-altering decisions (Bernhardt, 2014; Southworth & Mickelson, 2007). As a result, students of color are less likely to be placed in higher tracks, even with comparable academic records as their white counterparts, and certain groups of students are consistently disadvantaged because of vague criteria or preconceived biases (Bernhardt, 2014; Oakes, 1992).

Since the middle of the 20th century, tracking has been associated with efforts to maintain racial segregation in the public school system. After *Brown v. Board of Ed.*, many school districts used ability grouping structures to create de facto separate black and white learning environments. For this reason, courts began overseeing school systems and required school districts to take action in order to achieve unitary, or desegregated, status (Burris, 2014). These efforts have fallen far short of exorcising discriminatory tracking practices from the public schools (Watanabe, 2007). In 1992, for example, only about one-third of Latino and black students were enrolled in college prep tracks, compared to closer to 50 percent of white and Asian students. These figures have grown more disparate over time and lend credence to the idea that "the tracking system has had racial and ethnic overtones since its inception" (Hochschild & Scovronick, 2003, p. 161).

External pressures have also affected school tracking structures, particularly in mathematics. Following the launch of the Soviet satellite, *Sputnik*, in 1957 and fearing that Americans would lag behind in the international space and arms race, educational leaders called for major changes to mathematics education. The National Council for

Mathematics (NCTM) standards published in 1989 reflect the attention given to mathematics education, since the standards open with the declaration that "Mathematics has become a critical filter for employment and full participation in our society" (Standards, 1989, p. 4 as cited in Stinson, 2004, p. 10). Though NCTM's standards called for all students to be given the opportunity to become mathematically literate, ability grouping in mathematics became increasingly prevalent as the 20th century progressed. Consequently, students of color and female students were largely excluded from advanced mathematics courses (Stinson, 2004).

In 1997, the US Government published a white paper entitled "Mathematics Equals Opportunity" based on data from NELS: 88, which included 88 samples of 24,599 eighth graders from 1,052 schools, and the 1992 follow-up study of 12,053 students. The white paper presented data showing the importance of Algebra as a gateway subject to advanced math and science classes in high school. Also, the white paper stated that low income and minority students were significantly less likely to take higher level math classes, despite the importance of them doing so (Stinson, 2004).

As tracking practices have become more prevalent across America's schools, the persistent achievement gap between subgroups of students has also widened, with black and poor students performing consistently lower than their white, Asian and more affluent peers (Harris & Herrington, 2006; Reardon, 2011; Thernstrom & Thernstrom, 2003). Consequently, tracking research has focused on the correlational and causal relationship between ability grouping and gaps in achievement outcomes between racial subgroups. These efforts have led to a growing consensus around some of the factors tied

to tracking that contribute to the achievement gap. Nevertheless, scholars continue to disagree on the extent to which tracking contributes to educational inequities.

Tracking Research

Research on tracking practices first emerged in the educational literature around the middle of the 20th century. Since then, a great deal of research has established that tracking structures contribute to inequities in American public education. Nevertheless, questions remain about the exact nature of the relationship between tracking and disparities in achievement outcomes. In addition, research has not fully established how to address factors related to tracking that contribute to achievement disparities. The review of the literature that follows summarizes key studies and findings related to tracking structures across school levels.

Typically, researchers have studied tracking through two primary methods. In track/no track studies, researchers attempt to quantify the impact of tracking on student achievement by comparing students who are tracked with students who are not tracked. These studies can be either experimental or quasi-experimental in design (Burris, 2014; Slavin, 1990). In high track/low track studies, researchers compare the performance of students in higher-level courses with their peers who are enrolled in lower-level courses (Alexander & Cook, 1982; Burris, 2014). Many of these studies, particularly more recent studies, attempt to ascertain whether tracking affects students of different ability levels differently, as measured by prior achievement.

Many tracking studies focus on math achievement since, unlike language acquisition and reading skill, advancement in mathematics is more dependent on curriculum and teaching than on home environment or external factors. In addition,

mathematics' course placement in younger grades serves as a leading indicator of college preparatory course enrollment at the high school level (Domina & Saldana, 2012; Gamoran & Mare, 1989). Furthermore, there exists a wider achievement gap across socio-economic and racial lines in mathematics than in any other academic subject area (Boaler & Sengupta-Irving, 2016).

An extensive amount of research supports the notion that ability grouping expands achievement gaps between privileged and underprivileged groups (Gamoran & Mare, 1989). Early tracking research, from the mid-1950s, emphasized the ways that tracking practices promulgate a hierarchical social class order. In addition, researchers from the era examined how social class and a student's race affected track placement, particularly in higher-level classes. For example, sociologist Talcott Parsons' (1959) research on tracking, conducted in the late 1950s, demonstrated that schools typically operate under the supposition that advanced course offerings are scarce resources. Under this paradigm, it is impossible to provide all students with access to higher level and rigorous learning experiences. Schools address this notion of scarcity by limiting access to higher level courses according to measurements of merit or ability. Parsons' research also showed that curriculum differentiation, or tracking, played a key role in determining the future societal roles of children (Burris, 2014; Parsons, 1959).

Researchers that followed in Parsons' footsteps looked at the ways that factors such as parent levels of education, parent occupations and the number of books in the home might impact track placement. This research showed that social status more strongly correlates with track placement than do measures of student academic ability.

As a result of these research findings, researchers began questioning whether test scores should remain the primary factor in determining track placement (Burris, 2014).

In the 1970s, there began to emerge a growing realization among educational leaders and researchers that tracking practices tend to reinforce social class advantages that already exist in American society. Researchers who conducted studies during this time period found that tracking gave advantages to students from higher socioeconomic backgrounds, partly by giving those students more contact with higher-status peers (Burris, 2014). These studies underscored the notion that tracking practices reflect and reproduce the existing social hierarchy (Heyns, 1974).

Education researcher G.E. Hall confirmed that tracking exacerbated inequities in the American education system and society at large in a 1970 paper on inequality in America. Hall's paper explored several facets of track placement, including whether placement in lower ability classes resulted in a student having the ability to move up in tracks later on in his or her academic career. Hall also examined whether schools use fair and objective methods to determine track placement. Finally, Hall (1970) looked at student achievement data to determine whether tracking works as an educational strategy to improve student learning.

Legal decisions from this time period echoed these conclusions. For example, the 1967 federal court ruling in *Hobson v. Hansen* stated that Washington, DC schools had engaged in the de facto segregation of students by race through tracking placements. Specifically, the ruling highlighted the use of intelligence testing as being biased in favor of white and middle class students and required schools to move away from IQ test scores as the basis for track placement (Burris, 2014). Approximately ten years later, the

US Commission on Civil Rights published a report in support of the court's *Hobson* ruling. The Commission's report concluded that tracking, or ability grouping, was "the most common cause of classroom segregation," and that it led to racially segregated designations within schools (Burris, 2014, p. 8).

Janet Eyler's research on tracking practices in the 1980s built on the conclusions of previous literature. Eyler found that students were more likely to be placed in lower-level classes if they were poor or non-white. In addition, Eyler's research showed that track placements were rigid over time, with little movement between tracks. Eyler also found that students in lower-level classes spent less time on instruction than their counterparts in higher level classes. Finally, Eyler found little evidence to suggest that tracking provided targeted instruction for different ability groups. Instead, Eyler's research showed that students in lower tracks tend to fall farther behind over time and do not receive instructional supports that would enable them to develop into higher performing student (Burris, 2014; Eyler et al., 1982).

In 1985, Jeannie Oakes published *Keeping Track: How Schools Structure Inequality*, largely considered to be the definitive work regarding the drawbacks of tracking practices (Burris, 2014). Oakes' presented an extensive analysis of tracking practices and data from a diverse array of 25 junior and senior high schools located across the United States. 24 of the 25 schools had some type of tracking structure in place. In addition, all 13 high schools included in the study tracked students in English, math and science and nearly all of the junior high schools in the study tracked students in math and English (Oakes, 2005).

In Oakes' study, seven of the high schools and six of the middle schools were predominantly white, while eight schools were racially diverse. Two of the remaining four schools were predominantly black and two were predominantly Mexican American. Aggregate data collected from six of the racially diverse schools showed that 62 percent of students in high-track English classes were white, a disproportionately high number of students compared to the population as a whole. Conversely, only 29 percent of students in low-track English classes were white, a disproportionately low number of students compared to the population as a whole. In math, the same pattern existed, with the numbers being 60 percent and 37 percent, respectively (Oakes, 2005).

During the course of her research, Oakes identified the assumptions and beliefs held by supporters of tracking structures in schools. According to Oakes' research, educators who support tracking believe that students learn better when they are grouped with other students who are academically similar to them. Similarly, Oakes found that educators who support tracking believe that gifted students will not learn as much if they are placed in academically mixed classrooms and that slower students are better supported in lower level classes. Proponents of tracking also maintain that students of lower academic ability will develop stronger self-esteem if they are placed in classes that do not include students who are far more advanced than them. Finally, Oakes found that tracking supporters believed that placement processes and criteria were objective and fair (Oakes, 2005).

Oakes' research refuted many of these assumptions. According to the author, "mountains of research evidence" exists to prove that homogeneous grouping does not help anyone learn better (Oakes, 2005, p. 7). Oakes found that tracking created separate

and unequal learning environments for students. In lower-level classes, students learned less, were not as challenged and did not participate as much as students in higher-level classes (Oakes, 2005). As a result, students placed in lower-level classes developed lower self-esteem and self-efficacy than students placed in higher-ability groups. Similarly, Oakes found that peers and teachers believed that students in lower-level classes have less learning potential than students placed in higher-level classes (Oakes, 2005). By contrast, students in higher-level classes learned "how to do what scientists do," and learned how to conduct college-level research (Burris, 2014, p. 13). Oakes also found that strong racial patterns were associated with tracking at all levels. Even in vocational programs, minority students were more likely to be enrolled in lower-level courses. (Burris, 2014; Oakes, 2005).

Additional research conducted during the second half of the twentieth century and into the 21st century corroborated Oakes' findings. In the late 1970s and early 1980s, school districts in Illinois, California, Delaware and Pennsylvania were required to meet court-ordered unitary status requirements. Kevin Welner (2001), an educational researcher and legal expert, collected data related to these cases and presented his findings in a book titled *Legal Rights, Local Wrongs*. Welner found that tracked classes were not academically homogeneous and that, even when controlling for prior academic achievement, minority students were placed in lower-level classes at disproportionate rates. Like previous studies, Welner also found that track placement was generally rigid, meaning once a student was placed in a lower-level track, the student was unlikely to ever move up. Finally, Welner compared the academic outcomes of students who started at the same achievement level but were placed in different academic tracks. He found

that students who were enrolled in the lower-level track, experienced less academic growth than their counterparts who were enrolled in higher-level courses (Welner, 2001). A range of high track/low track studies, conducted around the same time period, affirmed these findings and concluded that students placed in higher tracks do better and that low-track classes have a negative effect on student achievement (Gamoran & Mare, 1989).

Throughout the 1980s and 1990s, Oakes, along with a team of RAND researchers continued studying educational tracking practices. Oakes and her co-authors found that teachers and students often perceived classes with a large number of minority students to be low-ability classes. Similarly, the authors found that higher percentages of minority and low-SES students were typically enrolled in less challenging math and science courses and schools with a high concentration of low-SES and minority students offered less high-level courses than more affluent and white schools. Finally, the authors demonstrated that teachers of lower-level courses were less experienced and qualified than teachers in higher level courses (Oakes, Ormseth, Bell & Camp, 1990).

Since the 1980s, researchers have also conducted studies in order to determine whether tracking is associated with gains or decreases in student achievement outcomes. Some of these studies divide students by subgroup based on ability levels, as measured by prior achievement metrics. Different studies reached different conclusions about the impact of tracking on student achievement for students of varying ability levels.

D. Veldman and J. Sanford's (1984) study on tracking found that students of lower ability are better off being placed in classes with higher achieving students (Burris, 2014; Veldman & Sanford, 1984). Veldman and Sanford's study included approximately 130 tracked junior high school classes in math and English. Using the California

Achievement Test to measure student ability, Veldman and Sanford found that both higher achievers and lower achievers did better in classes with a higher mean score. In addition, Veldman and Sanford showed that this effect was greater for lower achievers than higher achievers in both English and math classes (Veldman & Sanford, 1984). The authors also found that peer influences had a greater impact on students placed in lower level courses than on students placed in higher level classes. When lower-ability students were all grouped together, they were more likely to engage in teacher-dependent and off-task behaviors; however, when lower-ability students were placed in groups with higher achievers, they were more likely to adopt higher-achieving behaviors (Burris, 2014; Veldman & Sanford, 1984).

Alexander and Cook (1982) conducted a high track/low track study used Educational Testing Service (ETS) data collected from 1961-1969 to compare student performance across high track and low track classes. The study found that tracking had no significant impact on student achievement (Alexander & Cook, 1982). A few years after the publication of this report, Slavin conducted a track/no track study that confirmed this conclusion and found that the impact of tracking on student achievement was "indistinguishable from zero," (Slavin, 1990, p. 485 in Burris, 2014, p. 37). Slavin's and other researchers' findings meant that grouping students based on measures of prior performance had no significant effect on student achievement, when both groups were taught the same curriculum (Hochschild & Scovronick, 2003).

Slavin's findings slightly contradict an oft-cited earlier study by Kulik and Kulik (1982), that found that there was a slight increase in student achievement for higher ability students in tracked settings. Both Kulik and Kulik's and Slavin's studies,

however, did not find any statistically significant difference in achievement outcomes for lower and middle-level students in tracked or untracked settings (Burris, 2014; Kulik & Kulik, 1982). In 2009, University of Pennsylvania researcher Ning Rui affirmed Slavin's 1990 findings on tracking. Rui conducted a meta-analysis of 15 previous tracking studies, and found that placing students in heterogeneous groups or classes benefited low achievers and did not negatively impact the achievement of middle or higher level learners (Rui, 2009).

Mason et al. (1992) conducted a study of 34 average-achieving seventh-grade students who were placed in an advanced pre-algebra class. The authors found that the performance of the average achievers improved on standardized assessments. In addition, the study found that the average students who were placed in the advanced class experienced more growth in mathematics when compared with similar students who were not placed in advanced mathematics classes. Finally, the authors found that the average achievers who were placed in the pre-algebra class were more likely to enroll in higher level math classes in high school than other average achievers (Mason et al., 1992).

As a review of these studies demonstrates, scholars continue to disagree about the impact of tracking on achievement outcomes for students of different ability levels. Furthermore, the question of whether tracking negatively affects the achievement outcomes of higher ability students is a particularly sensitive topic that continues to gain traction in many school communities (Loveless, 2011). Thus, further research in this area is warranted, in order to continue to flesh out the specific impact of tracking on achievement outcomes when controlling for prior ability level.

Additional research in the 1990s focused on race and prior ability levels, to determine whether these factors are significantly associated with track placement. For example, some studies used regression analysis to show that prior achievement explained some, but not all, of stratification by race and class in higher course levels (Burris, 2014). Other studies, however, controlled for test-score performance and found that blacks were no more likely to be under-enrolled or over-enrolled in advanced math classes than whites. As a result of this uncertainty, Ronald Ferguson (1998), an education researcher at Harvard, noted that "the claim of racial discrimination in group placement by teachers is not supported by research, once conventional indicators of merit or economic standing are accounted for" (p. 329). This is an area that warrants further research, given the scholarly disagreements and the important nature of the questions involved.

Research has also been conducted to determine the correlation between socioeconomic status and placement in ability groups. Studies indicate that almost three times as many high-income students are placed in college preparatory tracks than low-income students. Researchers who have studied this phenomenon acknowledge that this disparity in placement can partly be attributed to the lower quality educational environments to which poor children are exposed (Hochschild & Scovronick, 2003).

Differences in parent involvement levels also helps to explain disparities in enrollment and in achievement outcomes. Elizabeth Useem studied the impact of parent involvement on track placement by looking at accelerated math class placement in middle school at 26 school districts in the greater Boston area. Useem found that parents with college and more advanced degrees were more likely to advocate for their children to be placed in accelerated math classes. Highly educated parents used their knowledge and

social capital to get their children into advanced math tracks. Conversely, less well-educated parents are more likely to trust the system and follow the educators' recommendations for math level placement (Useem, 1992).

In the early 2000s, Yonezawa, Wells and Serna (2002) conducted studies of schools where students and parents were permitted to choose their own course levels. In their study, the authors found that a range of factors contributed to the underrepresentation of black students and other subgroups from in higher level courses. Information regarding track choice was not evenly communicated among student groups. In addition, little to no effort was made to explain the importance of taking higher level courses to less well-educated parents. Minority students also complained that their desire to enroll in higher level courses was not taken seriously by counselors and teachers. Furthermore, black students were more hesitant to enroll in advanced classes because of peer pressure and the fear of being labeled as acting white. Black students talked about feeling supported in lower level classes, while feeling pressure to prove themselves in the upper level classes. The authors concluded that the only way to address these issues is to dismantle tracking structures entirely (Yonezawa, Wells & Serna, 2002).

Research conducted outside of the United States has also provided evidence that tracking structures contribute to disparities in achievement outcomes. Israeli researchers conducted study Project Together and Apart (TAP) to determine the impact of separating students by ability in mathematics. Researchers found that in a tracked setting, the gap between lower achievers and higher achievers expanded. In the untracked setting, where students were mixed heterogeneously, the gap between lower and higher achieving students did not expand. According to the TAP study, average and lower-achieving

students made significant gains in the heterogeneous classroom and higher-achieving students also made gains, though they were slightly smaller than the gains of high achievers in tracked classes. The researchers concluded that it was possible for all students to learn math in a heterogeneous, untracked setting (Linchevski & Kutscher, 1998).

Despite the findings of these and other research studies, proponents of tracking claim that grouping students by ability allows them to progress according to their ability, reduces failures, makes teaching easier and prevents bright students from being slowed down by less able peers (Oakes, 2005; Slavin, 1990). Advocates also believe that tracking tailors curriculum and instruction to specific student background experiences and prior achievements. In addition, supporters assert that schools must decide how to distribute limited resources to the most deserving students (Gamoran & Mare, 1989). Thus, tracking prepares students for the workplace, where resources are limited and where real-world competition exists based on ability, effort and interest (Bernhardt, 2014). In these ways, tracking supporters contend that tracking serves the greatest good for the greatest number of students (Gamoran & Mare, 1989).

Nonetheless, as this review of the literature highlights, the preponderance of research conducted over the past 75 years challenges these claims. Prior literature demonstrates that tracking structures have a negative impact on student achievement, particularly for mid- and lower level achievers. In addition, tracking structures reinforce the stratification of the American public education system by race and social class and reinforce the existing hierarchical social order (Southworth & Mickelson, 2007). Finally, tracking structures contribute to the achievement gap and that lower-level classes have a

negative impact on student achievement (Burris, 2014). These conclusions have formed the basis for school- and district-wide experimentation with eliminating tracking structures entirely by detracking or through universal acceleration policies.

Reducing the Achievement Gap by Detracking

For nearly four decades, educational experts and instructional leaders have experimented with and researched school structures that eliminate tracking as a way to build equity into the education system. Beginning in the 1980s, major civil rights groups, such as the NAACP, the American Civil Liberties Union (ACLU) and Children's Defense Fund all raised legal issues with tracking as a de facto second-generation segregation mechanism. The US Department of Education's Civil Rights Division also targeted tracking as an obstacle for compliance with Title VI regulations. The National Governors' Association, in 1989, proposed the elimination of tracking in order to meet national educational goals. Soon after, the National Education Association (NEA), the Council for Adolescent Development and other school reform bodies called for schools to abandon tracking practices in order to create caring, healthy, democratic and academically rigorous school environments. Federal court cases in the mid-1990s that targeted school districts in Pennsylvania, Illinois and California also found that tracking was racially discriminatory (Oakes, 2005).

Some states and school districts responded to these calls for policy reforms by moving more students into advanced-level classes, an effort known as curricular intensification. These school reforms focused primarily on mathematics instruction, since mathematics tracks are typically well-established across school districts and because school districts tended to accelerate students into more advanced mathematics

classes (Domina, McEachin, Penner & Penner, 2014; Domina & Saldana, 2012). Subsequent research on the impact of curricular intensification on student achievement outcomes is mixed and leaves room for additional analysis to determine the potential impact of these structural reforms on closing the achievement gap.

Research studies that took place from the late 1990s and early 2000s generally indicated that exposing students to more rigorous curriculum and instruction led to increases in student achievement (Argys, Rees & Brewer, 1996; Gamoran & Hannigan, 2000). As more school districts moved more students into advanced-level classes, however, the literature became more nuanced in terms of describing the benefits and drawbacks of these acceleration policies. For example, Domina et al., (2014) used a hierarchical linear modeling analysis to see whether student achievement outcomes changed when California began promoting a policy of all students taking Algebra in 8th grade. The study found that middle school mathematics curricular intensification had no effect on student achievement in small- and middle-sized districts. In large school districts, however, increases in eighth-grade Algebra enrollment corresponded with decreases in student achievement (Domina et al., 2014).

Discrepancies in the prior literature on the impact of curricular intensification on student achievement outcomes may be attributable to several factors. Firstly, large-scale intensification movements, particularly in larger school districts, can have a range of consequences on student achievement that may be difficult to measure or to capture precisely. For example, the quality of instruction, or the rigor of the curriculum in advanced courses might vary more drastically between classes and districts as the number of advanced-level course sections increase. Similarly, the level of teacher experience

may vary more broadly as the demand for advanced-level class teachers grows (Clotfelter et al., 2015). Secondly, studies may not fully account for selection bias, since students who are enrolled in advanced-level classes likely differ from peers enrolled in lower-level courses in a myriad of ways. As Domina et al. (2014) point out, relatively few studies have thus far attempted to address selection bias through an experimental or quasi-experimental design approach.

Like California, Chicago Public Schools moved toward universal acceleration in the late 1990s. This effort came to the fore in 1997, when Chicago Public Schools passed a policy mandating that all students enroll in Algebra I and English I (or higher), both college preparatory courses, by 9th grade. In their study of Chicago's policy change, Allensworth, Nomi, Montgomery and Lee (2009) focused on whether enrollment in ninth-grade college preparatory classes increased as a result of the new mandate. The researchers also looked at how changes in course level enrollment affected student achievement outcomes for students of different academic abilities (Allensworth et al., 2009).

Allensworth et al., (2009) found that Chicago's acceleration mandate led to a dramatic increase in college preparatory course enrollment, with the greatest impact occurring on low-ability student enrollment in advanced courses. The study also found that there was a ten percent increase in the number of students earning Algebra credit, including among lower ability groups. This figure can be partially attributed to the increase in the number of students taking Algebra. However, math failure rates for low-ability students also increased by about three percent. By contrast, in English, students at

all ability levels were much likely to earn English I credit and there were no adverse effects in terms of course failure rates.

Nevertheless, like previous studies in the area of course acceleration, this study's findings should be viewed in light of certain limitations. For example, Allensworth et al., acknowledge that variations between schools and teachers in terms of the quality of instruction, the rigor of the curriculum and the types of grading practices used in Algebra I classes serve as a limitation in this study. Consequently, the authors fail to arrive at a definitive conclusion in terms of the positive or negative impact of universal acceleration they end their study by calling for further research (Allensworth et al., 2009). This review of prior research, therefore, underscores the need for further study to try to determine the impact of accelerating students into more advanced math courses on student achievement outcomes.

While some schools and school districts have focused on the increased enrollment of underrepresented student subgroups in advanced course levels, other schools and school districts have eliminated tracking structures entirely. In 1999, for example, the Preuss School, a 6-12 charter school in California opened with the mission of preparing all of its students to be eligible to attend college. To achieve this goal, the school created an extended academic calendar and placed all of its students in the same challenging college prep track with additional supports as needed. The school also provided a personalized learning environment for each of its students by keeping enrollment and class size relatively small – 100 students per grade – and through a well-structured advisory program. Over 80 percent of the first graduating class attended a four-year college (Alvarez & Mehan, 2006).

The Rockville Centre School District in New York, embarked on a similar venture of detracking in the mid-1990s. The restructuring process occurred in response to the district superintendent setting a goal of ensuring that by the year 2000, at least 75 percent of all district graduates would earn a New York State Regents diploma. This number represented an increase of nearly 20 percent in the district and nearly 40 percent when compared to the state average (Burris & Welner, 2005). To accomplish this goal, the district universally accelerated all students, so that every student took the advanced math curriculum that had previously been reserved for only the highest performing pupils. In addition, in a radical move for the time, the district eliminated a "low-track" special education double period of mathematics, so that, beginning in 2001, the entire ninth-grade cohort of students with special needs was grouped in heterogeneous classes for all courses in the high school. These students were supported as needed with additional academic resources and all had access to the same pre-International Baccalaureate (IB) curriculum as their general education peers (Burris & Welner, 2005).

Prior to detracking, only 32 percent of African-American and Hispanic students in the Rockville Centre School District earned a New York State Regents Diploma, as compared to 88 percent of all white and Asian American graduates. After detracking was implemented, 82 percent of all black and Hispanic students met the criteria for a Regents Diploma as compared to 97 percent of white and Asian American students. Evidence shows that detracking in middle school and early high school years also had an impact on student access to advanced course opportunities later in their academic careers, since nearly half of all minority students were enrolled in IB English and History courses in eleventh grade, as compared to only 31 percent of minority students who were enrolled in

the same courses prior to tracking (Burris & Welner, 2005). In reflecting on the success of this new structure, the principal of the district's high school wrote that "achievement follows from opportunities," and that "the results of detracking in Rockville Centre are clear and compelling" (Burris & Welner, 2005, p. 598).

Despite the significant gains made in Rockville Centre, it is important to note that detracking alone did not solve the achievement gap issue, since schools modified instructional strategies and supports in order to meet the needs of all students in a heterogeneously mixed classroom (Alvarez & Mehan, 2006; Burris & Welner, 2005). Indeed, researchers agree that unless teaching methods are "systematically changed," maintaining or eliminating school grouping structures will have minimal impact on student achievement (Slavin, 1990, p. 491). Effective teaching strategies that must be implemented in conjunction with detracking structures include student-centered approaches to learning that promote a growth mindset philosophy among students and teachers (Boaler & Sengupta-Irving, 2016). Specifically, teachers must "problematize" content, give students voice and agency and hold students accountable to established disciplinary norms (Boaler & Sengupta-Irving, 2016). Similarly, cooperative learning methods provide a proven effective alternative to ability-grouping structures. This type of setting works best when students are able to establish group goals, when they are held individually accountable and when they work in small, heterogeneously-mixed groups (Slavin, 1990).

Findings from Boaler and Staples's 2008 study on tracking and mathematics student achievement reinforce Slavin's conclusions. Results from the study showed that students who attended Railside, an urban school with a diverse student body, made more

significant gains in mathematics achievement than the other two schools included in the study. At Railside, mathematics classes were non-tracked and teachers collaborated to provide all students with a rigorous curriculum that focused on conceptual understanding. According to the study, by senior year approximately 41 percent of all students at Railside were enrolled in advanced math classes, as compared to 27 percent of students at the other two schools. Railside also significantly decreased the achievement gap that existed between racial subgroups. In interviews, students stated that they learned to respect students who were different than them in their math classes and that they enjoyed the heterogeneous composition of their math classes (Boaler & Staples, 2008).

The San Francisco Unified School District (SFUSD) also recently implemented a district-wide reform to detrack middle school mathematics courses. Unlike other school districts, such as Rockville Centre, however, San Francisco adopted a policy that required all students to take grade-level 8th grade math, and did not promote universal acceleration of students into Algebra I. All students then take Algebra I in 9th grade and from there, students can decide whether to enroll in a compressed Algebra 2/Pre-Calculus class that prepares them to take Calculus prior to graduating. Though the policy is still in its infancy, and large gaps in research remain, early study data appear promising (Sawchuck, 2018). For example, more students are earning more math credits by the time they complete 11th grade, across all gender and ethnic subgroups. In addition, fewer students are repeating Algebra I and students across subgroups appear to be making gains in standardized test score results (Ryan, Barnes & Torres, 2018).

Like other areas of tracking research, scholars disagree about the impact of tracking reforms on narrowing the achievement gap (Oakes, 1992). For example, in the

mid-1990s, researchers used data from the National Educational Longitudinal Study of 1988 (NELS:88) to compare student achievement outcomes for students in tracked classes as compared to students enrolled in untracked classes. The researchers found that tracking caused a decrease in student achievement for students placed in lower-level classes and increased achievement for students placed in the higher track (Argys et al., 1996). The authors then attempted to estimate the effect of detracking on student achievement outcomes. The authors concluded that detracking would result in a nearly nine percent gain on math scores for students in the low track and approximately an eight percent decrease in scores for students in high-tracked classes (Argys et al., 1996). Despite limitations in the statistical analysis used in this study and later researchers' inability to replicate these findings, these numbers are often cited by proponents of tracking as evidence of the harmful impact that detracking would have on high-achieving students (Argys et al., 1996; Burris, 2014).

In his 1999 book, *The Tracking Wars*, Tom Loveless also asserted that detracking hurts high achievers (Loveless, 2011). In a follow up study, conducted ten years later, Loveless reviewed empirical research on the benefits and disadvantages of middle school tracking practices. Loveless' study examined data from the Massachusetts Comprehensive Assessment System (MCAS) to see if tracked and untracked schools produce a similar percentage of students reaching the advanced level. The author also reviewed survey response data related to tracking from 128 out of 295 Massachusetts middle schools. The results of Loveless' study indicate that tracking correlates to stronger student achievement. According to the study, more students score at the advanced level in mathematics in schools with three or more tracks and a reduction in the

number of tracks correlated to a three percent decrease in the number of advanced students (Loveless, 2009).

In addition to study findings, research has shown that opposition remains to the implementation of non-tracked schools because of beliefs and attitudes toward learning. For example, studies show that many people oppose detracking because they believe that minority students are not as capable learners as white students and that racially identifiable classes were simply the result of meritocratic criteria for placement (Oakes, 2005). Research has also shown that opponents of detracking efforts believe that intelligence is innate and is fixed at birth (Burris, 2014; Dweck, 2006). Finally, research has shown that opponents of detracking believe that school systems are meritocracies that are designed to reward the smartest and hardest working students and help create criteria to screen students for the best colleges. For this reason, families of high-track students in particular, typically oppose detracking efforts, because they generally believe that their students benefit from the tracking structures (Burris, 2014; Oakes, 1997).

Researchers must continue to explore how to modify existing tracking structures to best serve all students. Some researchers, for example, have suggested that tracked classes can be improved by raising the quality of lower-level classes, by using more objective course placement criteria and by providing equitable and heterogeneous opportunities outside of the classroom (Hallinan, 2004). In addition, studies have found improvements in student achievement outcomes in tracked schools when there is greater mobility between tracks and when more students are allowed to take higher-level courses (Oakes, 2005).

A review of the literature on tracking structures sheds some light on a path forward to tracking reform. For example, previous research shows that instructional strategies, such as cooperative learning and open-ended tasks can lead to improved achievement for all students, particularly when combined with efforts to reduce or eliminate tracking constructs. In addition, school districts must be willing to make changes and to stick with the change for long enough to collect evidence of the positive and negative effects of the change. School leaders must also be willing to engage with their communities in processes of professional learning and inquiry and to have open and honest conversations about tracking and its repercussions (Bernhardt, 2014; Watanabe, 2007). Finally, educational leaders must be willing to accept that there are high hurdles to overcome in order to detrack a school or school district completely, but that there are incremental steps that can be taken to address the social injustices and inequities that tracking produces (Oakes, 2005).

Further research is necessary and may continue to provide guidance to schools on tracking practices and acceleration policies. For example, additional research is warranted to learn how to replicate detracking success stories, such as those of Preuss and Rockville Centre across larger districts. In addition, future research must more fully examine the role that selection bias plays in student outcome and in prior study findings. Finally, future research should continue to explore the different effects that tracking and detracking has across different types of districts, including urban, suburban and rural districts and large- to small-size districts (Allensworth et al., 2009).

Ultimately, the conversation about tracking is important because the ongoing practice of dividing students by ability group reveals fundamental beliefs and norms of

the American education system (Oakes, 1992). Pre-determined course trajectories underscore the idea that each generation of students contains a distribution of ability that is roughly equivalent to the distribution of ability and effort of the previous generation and it is the purpose of the American education system to reproduce this status quo (Gamoran & Mare, 1989). "Ability grouping, like other forms of separation, has consistently provided the mechanism to give many students a second-class education" state researchers Hochschild and Scovronick (2003, p. 163).

Separating students into different tracks supports the notion that students arrive in school with an innate ability that is informed by their background and home life and that schooling can do little to change this pre-set course (Oakes, 1992). As a result, moving past the harmful impact that tracking has had on the American educational system will require a full normative and ideological shift, so that effective and equitable school structures are built on a solid foundation of equity-based principles (Trujillo, 2012). In this way, the American educational system may be transformed so that all students realize their full potential and have access to challenging and rigorous opportunity.

A Brief History of Shaker Heights

Cosmopolitan Magazine once called Shaker Heights the wealthiest city in the United States, and indeed, the city has a distinguished history in the annals of American suburbia (Meehan, 1963). The city takes its name from a colony of religious Shakers that had once lived in the area. First established as a village in February of 1912, Shaker Heights was already deemed to be "the finest residential district in the world," by a visiting aristocrat in 1924 (Molyneaux & Sackman, 1987, p. 20). In July of 1931, Shaker

Heights' residents adopted a charter and officially became a city (Molyneaux & Sackman, 1987).

Oris Paxton Van Sweringen and Mantis James Van Sweringen, two real estate mogul brothers who grew up in nearby Wooster, Ohio, are single handedly responsible for the creation of Shaker Heights. The Van Sweringen brothers never married, lived together until their deaths in the mid-1930s and are buried under the same tombstone. Throughout their lives, the two brothers were determined to realize their vision for an idyllic bedroom community outside of downtown Cleveland. In order to make their dream a reality, the Van Sweringen brothers purchased and tightly controlled the land that became Shaker Heights, bought a railroad to serve Shaker commuters and designed and built Terminal Tower, a landmark building in downtown Cleveland (Molyneaux & Sackman, 1987).

High quality public schooling was always a top priority for the Van Sweringens and other Shaker Heights' leaders. Less than a month after its official incorporation as a village, the Shaker Heights Board of Education asked voters to pass a $60,000 levy to build a new school. 20 out of 25 voters supported the bond issue and in September of 1912, the first Shaker Heights school opened with 26 students in attendance. Five years later, the Shaker Heights Board of Education instituted a policy requiring all Shaker Heights school teachers to have a college degree, a highly unusual measure for the time period (Molyneaux & Sackman, 1987). Shaker Heights' motto - "a community is known by the schools it keeps" - reflects Shaker's pride and emphasis in its school system (Ogbu, 2003).

In addition to excellent public schools, the Van Sweringens instituted strict building codes, requiring that all residential properties be built in Colonial, English and French architectural styles. The Van Sweringens also adopted a real estate covenant which gave their company the authority to approve or veto the purchase of any property lot within Shaker's borders. The Van Sweringen Compact, as it became known, was used largely to prevent black and Jewish families from moving to Shaker Heights.

As a result of these quality control measures, and the growing reputation of its school system, Shaker Heights property values rose and people began moving to the city. In 1911, the village census counted 200 inhabitants. By 1920, that number had increased to 1600 and within a decade, Shaker housed nearly 18,000 residents. In 1960, the number of Shaker residents peaked at approximately 36,400 (Molyneaux & Sackman, 1987). Today, Shaker Heights comprises approximately 27,500 residents (US Census Bureau, 2016).

In the late 1940s, the Supreme Court ruled that real estate covenants, such as the Van Sweringen Compact, were illegal, and over the next decade, the Compact largely ceased to be put to use. As a result, by the early 1950s, black families had started to move from Cleveland and other areas to the Ludlow area of Shaker Heights. This integration proceeded rather unremarkably, until one night in January of 1956, when a bomb destroyed the site of a new home being built by a black family in Shaker. The bombing spurred community-wide conversations between black and white families and ultimately led to the creation of the Ludlow Community Association in 1957. The Ludlow Community Association strove to promote a welcoming community for all, and

to assist both black and white homeowners to move into areas that were racially integrated (Molyneaux & Sackman, 1987).

Following *Brown v. Board of Ed.*, the Shaker schools also moved toward racial integration in the 1950s and 1960s. However, significant racial imbalances continued to exist in several of the district's elementary schools. Consequently, in the spring of 1970, the Shaker School Board adopted the Shaker Schools Plan, which offered voluntary busing to families in order to create a more racially balanced demographic across all Shaker schools. This plan remained in place until 1987, when the Board decided to close several elementary schools and to reorganize the secondary school structure in order to create more racial balance. As a result of this reorganization, all Shaker students now attend the same school from grades five through 12 (Molyneaux & Sackman, 1987; Pourdavood, Cowen, Svec, Skitzki & Grob, 1999). As one Shaker resident from the time put it, the Shaker community knew that "to get beyond racism, race must be taken into account" (Molyneaux & Sackman, 1987, p. 83).

The Van Sweringen brothers initially founded Shaker Heights in an attempt to create an exclusive country-club enclave, primarily for affluent white and Protestant families. The forces of integration, however, propelled Shaker Heights to the forefront of national conversations around equity and excellence in education. As a result, Shaker Heights has proven to be a fertile site for research on the interplay between race and academic achievement. The review of the literature that follows summarizes several key studies and findings that have emerged from research conducted in the Shaker Heights City School system.

Black-White GPA Disparities

In the spring of 1999, Ronald Ferguson and a team of Harvard researchers alighted in Shaker Heights to examine disparities in academic achievement between black and white students. At the time of the study, the mean Grade Point Average (GPA) for white students in Shaker was a full grade above those of black students, with white students having a mean GPA of a B+, and black students having a mean GPA of a C+ (Ferguson, Ludwig & Rich, 2001).

To study the factors that contributed to this achievement gap, Ferguson and his research team administered the Cornell Assessment of Secondary School Culture (CASCC) to all seventh through 11th grade Shaker Heights students at the end of the 1999 spring semester. A total of 1699 students responded to the survey, which included nearly all enrolled students. 83 percent of the respondents were either black or white (Ferguson et al., 2001).

The CASCC was designed to measure the impact of a range of factors on student academic achievement. Questions in the survey addressed a range of variables, such as race, gender, parents' years of schooling, household composition, social perceptions, and attitudes and behavior in school and out of school. For example, the survey asked students to explain what factors that might influence them to not study or to fail to complete homework. Students could select from a range of options such as competing commitments, being able to get a good grade without studying, carelessness, preferring to hang out with friends or believing that the work was too difficult. The survey also asked students about their attitudes and behaviors inside and outside the classroom, such as: the amount of time spent watching TV outside of school, aspirations to attend college,

whether the student studies with friends, whether the student copies assignments from friends, reasons why the student works hard and whether the student has friends who think that academic drive is not cool. Finally, the survey asked students about the characteristics of people whom they considered to be popular.

The authors' data showed discrepancies beyond academic achievement. According to the survey, approximately 90 percent of parents in white households had at least a college degree, while only about 45 percent of parents in black households had a college degree. In addition, parents had 12 or fewer years of schooling in 25 percent of black households, with the same being true for only five percent of white households. Furthermore, 52 percent of black males and 53 percent of black females lived with one or neither parent, with the same being true for 8.6 percent of white males and 14.6 percent of white females. Like other similar studies, the Ferguson study also showed that black students typically watched more TV at home than white students, though black students also spent more time on homework, in most cases, than comparable white students. Survey data also showed differences in how black and white students perceived what it meant to be popular.

Ferguson and his team conducted a multiple regression analysis to determine which factors most closely predicted or explained discrepancies in academic achievement. In addition, the Ferguson team controlled for factors related to family background and composition, in an attempt to determine whether a students' race was disproportionately lined to GPA disparities. The researchers found that a student's race was less statistically significant, when controlling for variables related to family background and socioeconomic status (Ferguson et al., 2001).

Ferguson et al. (2001), found that gaps in black-white student achievement were partly associated with variables related to family background. Study data showed that parents' education levels correlates with student academic achievement. Better-educated parents are associated with more academically engaged students who are better at studying, and who have a more stable school experience. A similar correlation exists for students living with two parents as well (Ferguson et al., 2001).

Nonetheless, Ferguson et al. (2001), found that "the single largest predictor of the black-white GPA gap is the proportion of courses taken at the honors and AP levels" (p. 372). According to the authors' analysis, parental education, household composition, student attitudes and behaviors are all factors that help explain why fewer black students take honors and AP classes than white students in Shaker Heights. However, the authors also stated that half of the differences in honors and AP class enrollment remained unexplained by these variable, which would seem to imply that racial bias may play a role in the enrollment disparity (Ferguson et al., 2001).

Beyond advanced course enrollment, Ferguson et al. (2001), found that effort, social pressures, behaviors and attitudes all have an impact on the student achievement gap. For example, blacks and whites spend approximately the same amount of time working on homework, but black students complete less homework than whites. The researchers noted that teachers only see the finished product, which may lead many teachers to have lower expectations for black students, since they believe that black students dedicate less time and effort to homework. Black students in Shaker Heights are also more likely than white students to see honors and AP classes as being socially and academically isolated, which may contribute to them being less likely to enroll in these

types of advanced classes. Furthermore, the study showed that students who do not take any honors or AP classes are more likely to claim that they do not work as hard as they are capable of due to peer pressure. Because black students in Shaker Heights are less likely than white students to take any advanced classes, they are also more likely to claim that they do not work as hard as they are capable of due to peer pressure. (Ferguson et al., 2001).

The researchers also found that the black-white gap in Shaker Heights was attributable to a gap in skills between the two racial subgroups. For example, on the Ohio Sixth Grade Proficiency Test in 1999, 91 percent of white males and 89 percent of white females passed the reading portion of the exam, as opposed to 51 percent and 41 percent of black males and females respectively. Math proficiency test results were similar. According to the authors, these standardized test results indicate that the average black student in Shaker Heights is simply less well prepared than the average white student to do well in honors and AP courses. Consequently, the researchers were not surprised by the disparities in black-white student enrollment in advanced level courses (Ferguson et al., 2001).

As a result of these and other factors, Ferguson et al., asserted that white students in Shaker Heights typically have a head start when compared to black students, in terms of educational content knowledge and also in terms of knowledge about how the educational system works. The authors also note that most students who are not enrolled in honors or AP classes reported having a lower GPA than those students who took advanced classes. Consequently Ferguson et al., discouraged Shaker Heights from pushing students "wholesale" into honors and AP classes (Ferguson et al., 2001, p. 373).

Based on their findings, the authors made several recommendations to attempt to narrow the achievement gap. First, the authors stated that increased effort on the part of black students might contribute to increases in black student achievement. In addition, the authors recommended the implementation of instructional strategies designed to improve black student engagement. These types of instructional practices include clearly defined purposes and goals and provide students with strategies to use apply in order to meet the goals. Students should also be provided with rewards that motivate them to accomplish academic goals (Ferguson et al., 2001).

Ferguson et al., did acknowledge several limitations to their study. Because of the nature of the survey, the study relied on student self-reported data. In addition, the authors conceded that it was difficult to distinguish between causal and correlational relationships among the data. Finally, the authors did not explore grade-level differences in the data (Ferguson et al., 2001).

Academic Disengagement

In the spring of 1997, a group of black Shaker Heights families reached out to John Ogbu, a Nigerian-American anthropologist known for his work on race and academic achievement. The families contacted Ogbu because they were increasingly concerned by the black-white achievement gap in Shaker Heights and their concerns had been heightened by a recent Shaker Heights High School newspaper article calling attention to the problem. Ogbu and his team agreed to conduct an ethnographic study of Shaker Heights, a research project that was ultimately funded by the school district and that was conducted with the full support and cooperation of the community (Ogbu, 2003).

Ogbu's research focused on some of the reasons behind the low performance of black students in the Shaker Heights schools. As part of examining this research problem, Ogbu explored why black students in Shaker Heights were disengaged from their academic work as compared to white students. The research primarily focused on societal, school and community forces that contributed to these issues (Ogbu, 2003).

To answer these research problems, Ogbu conducted a qualitative ethnographic research study of the Shaker Heights School System. The study lasted for eight months, half of which were spent doing continuous fieldwork. The research team visited and observed instructional practices and interactions in three of Shaker's elementary schools and all of Shaker's secondary schools and collected data through group and individual discussions, individual interviews, archival documents and participant observation. In his published findings, Ogbu dedicated ample space to the participants' own words and reflections. In this way, Ogbu's study provided a platform for the voices of the Shaker Heights community to explain and uncover some of the root issues that continue to challenge the school community (Ogbu, 2003).

Data that Ogbu collected show that in the 1995-1996 school year, blacks performed worse than whites in every measure of academic performance. This achievement gap was visible on assessment metrics as well as in person. "In almost every school we visited," wrote Ogbu, "there was some direct or indirect evidence of racial differences of the performance status" (Ogbu, 2003, p. 7). Indirect evidence included the number of black students enrolled in higher level courses, particulate at the upper elementary school and above (Ogbu, 2003).

Ogbu's research pointed to several underlying issues as the cause of the persistent achievement gap. Generally speaking, Ogbu found that black students did not work as hard as white students. Ogbu ascribed this lack of hard work to several factors, some of which were outside the immediate control of students or families. For example, Ogbu noted that the long history of slavery and discrimination has inculcated the black community with a set of challenges in terms of fully integrating into a school system that is largely seen as driven by white culture. Furthermore, black students struggled to connect academic achievement with long-term career success. Finally, black students were less likely to have professional middle-class role models who encouraged them to pursue careers that depended on educational, rather than athletic, success (Ogbu, 2003).

Despite these findings, Ogbu also found that there were factors that contributed to weak black academic achievement that were more directly in the control of students and their families. For example, parents of black students were less likely to be directly involved in their child's education, as measured by attendance at parent/teacher conferences and other school events, and were therefore less likely to be fully informed about the impact that taking advanced courses has on a student's overall academic achievement and future career prospects. In addition, Ogbu found that black students sometimes shied away from working as hard as their white peers, because they did not want to be perceived as acting white. Black students were also less likely to study at home and engaged more frequently in competing distractions, such as TV watching or athletics, than their white counterparts. Finally, black students were more likely to be disengaged in class and have more discipline problems than white students, an issue

which Ogbu and his team acknowledged might be partially attributable to low teacher expectations for black students (Ogbu, 2003).

According to Ogbu's study, school structures also contribute to Shaker's achievement gap. In this instance, Ogbu focused primarily on the role of course leveling, or tracking on student achievement. Leveling is problematic, because students are sorted into different academic pathways early on in their academic careers and these pathways are unlikely to change significantly once they are established (Ogbu, 2003). In Shaker Heights, these pathways largely differ based on the race or ethnicity of the students, with black students dominating lower level courses and white students largely enrolled in upper level classes. Ogbu found that students who are placed in lower level courses in Shaker Heights are exposed to lower level instruction that focuses more on rote tasks, such as memorization of facts, rather than on comprehension and analysis. Finally, because of the segregated nature of these classes, black students were less likely to feel comfortable in advanced classes, with the same being true of white students in grade-level classes (Ogbu, 2003).

Ogbu presented several possible solutions to improve the segregated nature of the Shaker School district and to address the significant achievement gap. He advocated for the establishment of extra- and co-curricular activities that provide black students with mentors and role models who demonstrate the importance of educational success. According to Ogbu, the Minority Achievement Committee (MAC) Scholars program in Shaker Heights, a club for high achieving black students, exemplified this type of approach. Through the MAC Scholars program, black students meet other academically successful black students and receive mentorship, tutoring and a consistent message that

academic achievement is attainable and preferable to other types of success (Ogbu, 2003).

Ogbu also called for the black community to rally around its students and to become more involved in the school district, so that they can become more aware of the steps that can be taken at home to bolster academic achievement. "The black community and black families must assume a proactive role to increase the academic orientation, effort, and performance of their children" wrote Ogbu (Ogbu, 2003, p. 274). To achieve this goal, Ogbu encouraged black family members to teach their children how to work hard and make good grades, how to avoid distractions at home and to communicate the importance of education in the context of future career success (Ogbu, 2003).

Ogbu's final recommendations centered on initiatives that the school district can undertake to promote equity of academic achievement. For example, Ogbu encouraged school leaders to promote student-centered and collaborative instructional practices. While Ogbu stopped short of calling for the elimination or eradication of course tracking, he did implore the school district to do everything in its power to educate students and their parents about course leveling so that they can make informed and effective decisions (Ogbu, 2003). Since Ogbu's research study, the Shaker Heights school district adopted an "open enrollment" policy to allow students to opt into advanced classes, even if they do not meet the testing or teacher recommendation criteria (Ogbu, 2003).

Sites of Educational Privilege

Cleveland State University professor and Shaker resident Anne Galletta also conducted a qualitative research study of the Shaker Heights school system. Like Ogbu, Galletta was interested in the intersection between educational achievement and race in

Shaker Heights. As part of her study, Galletta explored times when Shaker Heights' residents worked to create a more integrated community by reducing racial exclusion in the school system (Galletta, 2013).

Galletta's research looked specifically at how Shaker parents, students and teachers across race and class lines experienced racial equality, in the context of school desegregation efforts. Galletta defined equality as the: "providing of equal educational opportunities, producing equal academic outcomes, and engendering equal power relations between students of color and white students" (Galletta, 2013, p. 16). Galletta examined educational structures, such as school policies and practices, that promoted racial equality or that reinforced societal inequities (Galletta, 2013).

Galletta framed her research in the critical theory interpretive paradigm and used a case study approach. Galletta's case study focused on the Shaker Heights school system and combined archival study and oral histories with semi-structured interviews. The study included extensive archival research and oral histories with 22 individuals identified as key players in the Shaker Heights community. Galletta also conducted semi-structured interviews with 43 participants. Through these methods, Galletta explored individuals' experiences in the Shaker Heights school system and the context within which they occurred. Galletta summarized her study as an analysis that relied on "the *experience* as narrated by participants within different opportunity structures" (Galletta, 2013, p. 19).

In her findings, Galletta focused on the competing forces that both opened up and restricted black student access to "sites of educational privilege," such as advanced courses (Galletta, 2013, p. 171). Galletta found that school-wide reforms, such as the

1970 Shaker Schools Plan to voluntarily bus families and the 1987 consolidation of Shaker schools, had helped removed barriers to racial integration. Other structural changes in Shaker, such as improvements in instructional practices and the MAC Scholars program, were also created with the stated goal of helping black students gain access to high quality educational opportunities over time.

Nevertheless, Galletta found that these efforts to promote equality engendered opposition that resulted in the erection of barriers to sites of educational privilege. For example, policies were put in place to assure white families that Shaker's high academic standards would not be diminished. Most prominently, Galletta found that tracking practices were used to maintain de facto segregated learning environments (Galletta, 2013).

Galletta concluded her research by noting that both black and white students in Shaker experience the tension between desegregation and segregation. Some black students in Shaker engage in sites of educational privilege, but often experience racial isolation that leads them to question their sense of belonging. White Shaker students, by contrast, deal with the expectation that they are supposed to be in advanced classes. Though Galletta did not find that this pressure led to any decreases in academic achievement for white students, she did note that both black and white students found advanced courses to be places of tension that were not race-neutral. Students' experiences navigating tracking structures led many students to confront racial stereotypes and to reconsider their own identities (Galletta, 2013).

Summary

As the preceding review of the literature shows, a substantial body of research has been dedicated to exploring the causes and effects of the black-white achievement gap. In addition, scholars have examined the impact of tracking structures on student achievement. Research that has been conducted in Shaker Heights provide case study and survey data that further the empirical understanding of black and white differences in educational experiences and outcomes. Because of Shaker's unique history and school structures, the school district remains a fertile ground to conduct research related to the interaction between academic pathways, race and student achievement.

The review of the literature provides insight into factors to consider in this research study as well as gaps that exist in current understanding. This study endeavors to strengthen the existing body of literature by further quantifying the impact that tracking has within and between racial subgroups in Shaker Heights. Furthermore, this study will address specific gaps in the current body of literature by focusing on how tracking affects the achievement outcomes of students of comparable ability levels. Finally, by examining discrepancies in student achievement outcomes, this study will provide additional evidence on how tracking in mathematics may affect the black-white achievement gap.

CHAPTER III

METHODOLOGY

Introduction

This study compared the achievement outcomes of students enrolled in advanced mathematics classes at the middle grades level, with students of comparable ability and background who are enrolled in grade-level math classes. Achievement outcomes for this study were measured using benchmark data from the NWEA MAP assessment in mathematics. As stated above, research questions for this study included the following:

1. How do the achievement outcomes of students enrolled in advanced math classes compare to the achievement outcomes of students with comparable ability levels and background characteristics who are enrolled in grade-level math classes?

   a. Over the course of one academic year, based on MAP assessment data, do students in advanced math classes achieve more or less than statistically similar students who are enrolled in grade-level courses?

2. How do the achievement outcomes of black students who are enrolled in advanced math classes compare to the achievement outcomes of black

3. Students with comparable ability levels and background characteristics who are enrolled in grade-level classes?

   a. Over the course of one academic year, based on MAP assessment data, do black students in advanced math classes achieve more or less than statistically similar black students who are enrolled in grade-level courses?

4. How do the achievement outcomes of white students who are enrolled in advanced math classes compare to the achievement outcomes of white students with comparable ability levels and background characteristics who are enrolled in grade-level classes?

   a. Over the course of one academic year, based on MAP assessment data, do white students in advanced math classes achieve more or less than statistically similar white students who are enrolled in grade-level courses?

Site and Program

The site of this study is the Shaker Heights City School District. This site was chosen because the school district is uniquely suited for a study of this type. Shaker Heights has been the focus of prior research related to the intersection between race, privilege and academic achievement. Historically, Shaker Heights City School District administrators and community members have been open to allowing researchers to study the district's students and programs, and this attitude continues today (Ogbu, 2003). In addition, the school district's structure, demographic composition, challenges and successes typify many aspects of school districts across the country; results from this study may therefore be generalizable to school districts outside of this study site. Finally,

as reviewed above, this site was chosen due to convenience, since the author serves as an administrator in the school district.

The author's employment in this school district serves to strengthen the study findings, because the author is deeply familiar with the school district's structure, program, curriculum and operations. Nevertheless, the author's position in this district can be considered a limitation of the study as well. Care was taken to identify and address any influence that the author's positionality may have on methodology and study design.

The sample for this study consists of data from students enrolled in Shaker Heights Middle School during the 2016-2017 or 2017-2018 school years. These years were selected, because Shaker Heights Middle School administered the MAP assessment in mathematics to all students during this time. Prior to the 2016-2017 school year, the MAP assessment was not administered consistently across the school. Permission to collect data was granted from the Cleveland State University Institutional Review Board. Shaker Heights Middle School is the only middle-years level school in the Shaker Heights City School District, which is a first ring suburb located just outside of Cleveland, Ohio. During the 2016-2017 school year Shaker Heights City schools served approximately 5021 students (Ohio Department of Education, 2017). During the 2017-2018 school year, Shaker school enrollment remained relatively constant (Shaker Heights City School District, 2018). The Shaker Heights City School District includes four Pre-K-4 elementary schools, one upper elementary school, which consists of grades five through six, one middle school, comprised of grades seven and eight, and one

comprehensive high school. The entire district is authorized as an International

Baccalaureate (IB) school district.

Enrollment during the 2016-2017 school year for Shaker Heights Middle School

was approximately 786 students, with 407 students enrolled in grade seven and 379

students enrolled in grade eight (Shaker Heights City School District, 2017). During the

2017-2018 school year, enrollment at Shaker Heights Middle School dropped to

approximately 760 students, with 367 students enrolled in grade seven and 393 students

enrolled in grade eight (Shaker Heights City School District, 2018).

As a whole, Shaker Heights City School District is divided nearly evenly between

black and white students. During the 2016-2017 school year, for example, the Shaker

Heights student body demographic was comprised of the following racial subgroups: 44.9

percent black; 41.1 percent white; 14.1 percent Asian, Hispanic or Multiracial.

Approximately 1520 students, or just over 30.0 percent, qualified for free or reduced

meals (FRM) during the 2016-2017 school year. 15.1 percent of students in Shaker

Heights were identified as having a disability and two percent of students were identified

as English Language Learners. Shaker Heights had an overall attendance rate of 96.1

percent during the 2016-2017 school year (Ohio Department of Education, 2017).

The demographic data of Shaker Heights Middle School mirror those of the

school district. During the 2016-2017 school year, Shaker Heights Middle School's

student body demographics consisted of the following racial subgroups: 46.8 percent

black; 40.2 percent white; 13.0 percent Asian, Hispanic or Multiracial. Just under 30

percent of middle school students qualified for FRM during the 2016-2017 school year

and 14.9 percent of students in Shaker Heights Middle School were identified as having a

disability. Shaker Heights Middle School had an overall attendance rate of 96.0 percent during the 2016-2017 school year (Ohio Department of Education, 2017). These numbers stayed relatively constant during the 2017-2018 school year.

The vast majority of students who qualify for FRM in Shaker Heights are black. In the fall of the 2017-2018 school, for example, over 85.0 percent of students who qualified for FRM in the district were black, and just 4.5 percent of students who qualified for FRM who were white. As this statistic reflects, there is a substantial socioeconomic racial disparity among the district's student body (Brazer, 2018).

During the 2017-2018 school year, the Shaker Heights school district commissioned a qualitative study of gifted programming in the school district led by Stanford University education researcher David Brazer. Brazer (2018) found strong evidence that a black-white test score gap exists in Shaker Heights. According to third- and fourth-grade MAP reading scores, only 1.2 percent of white students fall in the lowest quartile, as compared to 9.8 percent of black students. By contrast, nearly 40 percent of white students score above the 50th percentile in third- and fourth-grade MAP reading scores, as compared to just 20.0 percent of black students (Brazer, 2018).

At the end of the 2016-2017 school year, according to MAP score projections, 66.0 percent of seventh- and eighth-grade white students were expected to be on track to earn a 24 or higher on the ACT math exam, which is a leading indicator that they will be college ready in mathematics (Thum & Matta, 2015). Similarly, based on their MAP assessment scores, 71.4 percent of seventh- and eighth-grade white students were considered to be on track to score accelerated or advanced on the Ohio State American Institute of Research (AIR) end-of-course mathematics exam in the spring of that

academic year (NWEA, 2016). In the winter of 2017-2018, these numbers remained relatively similar, with 70.3 percent of seventh- and eighth-grade white students projected to be on track to earn a 24 or higher on the ACT math exam and 72.6 percent on track to score accelerated or advanced on the AIR end-of-course mathematics exam, based on the students' MAP assessment scores.

Black student achievement outcomes differ starkly from their white counterparts. In the spring of the 2016-2017 school year, 84.0 percent of seventh- and eighth-grade black students were not considered to be on track to earn a 22 or higher on the ACT exam in mathematics, according to MAP score projections. Where a projected ACT score of 24 demonstrates that a student is college ready, a projected ACT score of 22 is the minimum score necessary to indicate that a student has demonstrated proficiency in high school mathematics standards by the time they graduate. In the spring of 2017, only 9.1 percent of seventh- and eighth-grade Shaker black students were projected to earn a 24 or higher on the ACT exam in mathematics based on their MAP assessment scores (Thum & Matta, 2015). Similarly, only 10.6 percent of seventh- and eighth-grade Shaker black students were considered to be on track to score accelerated or advanced on the AIR end of course mathematics exam that spring (NWEA, 2016). In the winter of the 2017-2018 school year, 83.0 percent of seventh- and eighth-grade black students were not considered to be on track to earn a 22 or higher on the ACT exam in mathematics and only 13.0 percent of black students were on track to score accelerated or advanced on the AIR end of course mathematics exam.

Black students fared slightly better in terms of reading performance during the 2016-2017 school year. At the end of the 2016-2017 school year, MAP assessment

scores show that 74.0 percent of seventh- and eighth-grade white students were projected to be on track to earn a 24 or higher on the ACT reading exam. This figure compared with 21.7 percent of seventh- and eighth-grade black students, who were projected to score a 24 or higher on the ACT reading exam (Thum & Matta, 2015). These figures stayed relatively constant during the 2017-2018 school year, when 77.6 percent of white students were projected to be on track to earn a 24 or higher on the ACT reading exam, compared with 21.0 percent of black students.

Shaker Heights Middle School's state report card data correlate with the MAP data. On the 2016-2017 Ohio State School Report Card, Shaker Heights Middle School earned a score of an F in the category of Gap Closing. The metric compares the performance of racial subgroups toward meeting a state annual measurable objective (AMO). The 2016-2017 Ohio State AMO was 77.1 percent in English Language Arts and 72.0 percent in mathematics, which meant that every racial subgroup with more than 30 students was expected to meet this mark. In English Language Arts, 85.4 percent of white students at Shaker Heights Middle School met the AMO, compared with just 30.6 percent of black students. In mathematics, 82.3 percent of Shaker Heights Middle School white students met the AMO, compared with 33.1 percent of black students (Ohio Department of Education, 2017).

Ogbu (2003) noted similar racial discrepancies in academic achievement in Shaker Heights. In 1995, white sixth-grade students in Shaker scored 86 percent proficient in mathematics and 97 percent proficient in reading, whereas black sixth-grade students in Shaker scored 28 percent in mathematics and 70 percent proficient in reading. White eighth-grade students scored 92 percent in mathematics and 100 percent proficient

in reading, compared to 37 percent and 83 percent, respectively, for black eighth-grade students (Ogbu, 2003, p. 5).

A similar racial disparity exists in Shaker Heights in terms of the composition of advanced courses in Shaker Heights. In grades 5 through 12, nearly 65 percent of students in advanced classes are white, as compared to just 18 percent black. It is worth noting that nearly 70 percent of black students who took advanced classes in grades five through 12 in 2017-2018 in Shaker Heights qualified for FRM, as compared to just under seven percent of white students enrolled in advanced classes (Brazer, 2018).

This racial disparity in advanced course enrollment has existed in Shaker Heights for a substantial amount of time. During the 1993-1994 school year, over 60 percent of students enrolled in grade-level courses were black, compared to just 36 percent of students in grade-level courses who were white. By contrast, blacks comprised just over 20 percent of students enrolled in advanced placement and honors courses compared to 74 percent for white students (Ogbu, 2003).

Brazer's (2018) study also looked at how students are enrolled in advanced mathematics courses in Shaker. Brazer determined that the identification of students for advanced level classes and the implementation of these enrichment opportunities is inconsistent across schools. For example, some elementary schools in Shaker Heights adhere closely to published criteria related to student test performance in order to determine student course placement. In other elementary schools within the district, however, parent advocacy may hold greater influence in terms of student placement in advanced courses. In addition, in some elementary schools, an enriched, or advanced, mathematics curriculum is provided when specialized teachers push-in to work with

students in their regular math classes. In other elementary schools within the district, however, students are pulled out of class to receive advanced programming. Brazer concluded that these inconsistencies have led to confusion throughout the Shaker Heights school system (Brazer, 2018).

According to the district's Academic Planning Guide (Shaker Heights City School District, 2018), the district adheres to an "open enrollment" policy. At the middle school level, this open enrollment policy is defined as follows: "The Shaker Heights City School District has an open enrollment policy for all students interested in participating in any given course. Students are encouraged to pursue the highest level of instruction matching their motivation, interest, and previous learning" (p. 32). In practice, this policy means that a student who wishes to enroll in advanced courses may do so, even if he or she does not meet the prerequisite criteria for entrance into the course. Some schools in Shaker grant waivers to students who enroll in courses above their deemed proficiency level, while other schools simply admit the students into the advanced class (Brazer, 2018).

The open enrollment policy was created in an attempt to diversify advanced-level courses. Specifically, the policy was formed in order to spur an increase in the number of black students taking advanced courses and to decrease the disproportionate ratio of white students to black students in advanced level classes. The policy, however, has had the opposite of its intended effect. In 2016-2017, for example, 18 white students in sixth grade requested placement in Enriched Mathematics as compared to only six black students. In 2017-2018, the numbers are eight and two, respectively. These figures demonstrate that the open enrollment policy has reinforced patterns of racial disparity in advanced level classes, rather than eliminate them (Brazer, 2018).

Study Sample

The study sample consists of all Shaker Heights Middle School students who were enrolled for the duration of the 2016-2017 or 2017-2018 school years, as measured by attendance records and evidence of having taken all three NWEA MAP math benchmark assessments, in the fall, winter and spring of the school year. In addition, students were included in this study if they took a grade-level, advanced- or accelerated-level math class. Students who were enrolled in self-contained resource room or multi-handicapped math classes were not included in this study, because the subgroup was too small to be considered statistically significant.

This study included over 1,500 participants, with approximately 750 participants enrolled in Shaker Middle School during the 2016-2017 and 2017-2018 school years. Though many of the same students were enrolled at Shaker Heights Middle School for both years, each year will be treated as a separate case in this study. In addition, grade level enrollment was fairly evenly distributed between 7th and 8th grades during the years of the study (Shaker Heights City School District, 2018).

Measures

Data for this study were compiled from district administrative records.

Math achievement. The outcome variable for this study was student percentile scores in MAP mathematics during the current academic year. Shaker Heights City School District administers the MAP assessments in reading and mathematics to all seventh and eighth-grade students enrolled in general education classes three times a year, in the fall, winter and spring. The MAP assessments benchmark student progress

and are used to screen for enrollment in advanced courses and for other intervention support services.

The MAP assessment system measures achievement in reading, language usage and mathematics for grades two through 12. All test items, on both reading and mathematics tests, are multiple-choice format and are administered using a computer adaptive method. The test content aligns with Ohio State Standards in reading and mathematics, which are based on the Common Core standards (NWEA, 2011).

The MAP assessments quantify student ability levels and growth using a measurement scale based on the Rasch model. MAP scores are quantified using the Rasch unit (RIT) scale. The RIT scale was developed by NWEA for all MAP assessments. The RIT scale structure enables the MAP assessment to administer different test items to different students and to obtain comparable results. The RIT score structure allows the NWEA MAP assessment to be computer adaptive and targeted (NWEA, 2011).

NWEA conducts field testing to ensure that the MAP assessment in mathematics is both reliable and valid. Each test item is administered to a sample size of at least 1,000 students. Field testing results indicate that a student taking the same test, or a test of equal difficulty and similar content, multiple times, would receive the same score, across administrations. In addition, field testing results demonstrate that test items contain accurate content and information and that test items appropriately assess the topic and skill (NWEA, 2011).

According to Thum and Hauser (2015), the mean RIT score for a student at the seventh-grade level in 2015 was a 222.65, with a standard deviation of 16.59. The mean

for grade seven fall-to-spring growth for a student was 5.95, with a standard deviation of 6.55. The mean RIT score for a student at the eighth-grade level was a 226.30, with a standard deviation of 17.85. The mean grade eight fall-to-spring growth for a student was 4.63, with a standard deviation of 7.66.

A student who receives an overall RIT score of 223 in mathematics in the fall of seventh grade is performing at the 50th percentile level. For eighth grade, the RIT score number would increase to 226. A student who achieves a RIT score of 272 in the seventh-grade fall administration in mathematics is performing at the 99th percentile, with a RIT score of 276 being the corresponding score for an eighth-grade student. For the spring administration, a seventh-grade RIT mathematics score of 229 would fall at the 50th percentile, with a RIT score of 231 for the eighth grade. A seventh-grade spring RIT mathematics score of 275 would fall at the 99th percentile, with a RIT score of 281 for an eighth-grade student (Thum & Hauser, 2015).

In the area of reading, a seventh-grade fall RIT reading score of 214 would fall in the 50th percentile, with a score of 217 being the corresponding score for eighth grade. A seventh-grade fall RIT reading score of 250 in reading would fall in the 99th percentile, with the RIT score increasing to 254 in the eighth grade. A seventh-grade spring RIT reading score of 218 would fall in the 50th percentile and an eighth-grade student would need a score of 220 to fall in the same percentile. A seventh-grade spring RIT reading score of 253 would fall in the 99th percentile, with that number increasing to 257 for eighth-grade students (Thum & Hauser, 2015).

For purposes of this study, students were categorized according to their prior math ability level, as having low mathematics ability, average ability, or high ability. Prior

math ability level was calculated by taking the average of a student's three prior-year MAP mathematics percentile scores. Students with an average prior-year MAP mathematics percentile score of less than the 25th percentile were deemed low math ability; students with an average prior-year MAP mathematics percentile score that was greater than or equal to the 25th percentile and less than or equal to the 75th percentile were deemed average; students with an average prior-year MAP mathematics percentile score of over the 75th percentile were labeled as having a high mathematics ability.

Attendance rate was calculated by dividing the total days present by the total days enrolled for that school year. An attendance rate of one therefore, indicates that a student had perfect attendance during the period enrolled.

Tracking. The primary independent or treatment variable in this study is whether a student is enrolled in a grade-level versus an advanced-level math course. Shaker Heights Middle School offers several different levels of mathematics courses. A seventh-grade student is in an advanced math course if the student is enrolled in Pre-Algebra. A seventh-grade student is in an accelerated math course if the student is enrolled in Algebra I. An eighth-grade student is in an advanced math course if the student is enrolled in Algebra I. An eighth-grade student is in an accelerated math course if the student is enrolled in 9 Honors Math. Due to their small number, students who were enrolled in accelerated courses were not included in the final study. Thus, for this study, Shaker Heights Middle School students were included in the treatment course if they took advanced-level math during the 2016-2017 school year or 2017-2018 school year.

According to the 2016-2017 Shaker Heights Middle School programming placement guidelines, seventh-grade students were recommended to be enrolled in Pre-Algebra for the 2016-2017 academic year if they met the following criteria:

- Scored an overall MAP RIT score of 235 or higher on the winter MAP mathematics assessment

- Completed sixth-grade advanced level mathematics with a grade of an A or a B

- Were considered to be self-disciplined and motivated; cooperative when working in groups; able to apply understanding in new situations; attends to precision

- Demonstrated mastery in operations with fractions, relationships between percents, decimals and fractions, concept of area of perimeter (circumference) of polygons and circles, coordinate graphing and fluency with rational numbers.

These criteria remain relatively constant for the 2017-2018 school year. However, the overall MAP RIT score was changed to 231, instead of 235. All other criteria remained the same.

According to the 2016-2017 Shaker Heights Middle School programming placement guidelines, eighth-grade students were recommended to be enrolled in Algebra I for the 2016-2017 academic year if they met the following criteria:

- Earned an overall MAP RIT score of 239 or higher on the winter MAP mathematics assessment

- Completed seventh-grade Pre-Algebra course with a grade of an A, B or C

- Passed the seventh-grade Pre-Algebra midterm exam with a grade of a C or higher
- Were considered to be self-disciplined and motivated; cooperative when working in groups; able to apply understanding in new situations; attends to precision
- Demonstrated mastery in concepts required for Math 8 and demonstrated ability to represent linear relationships in multiple representations, understanding of proportionality, fluency with equivalent expressions, and operations with rational numbers including integers.

The criteria for 2017-2018 Algebra I placement remained the same.

Matching Variables

The following variables served as the basis for student matching in order to allow for equitable comparisons between the treatment and non-treatment groups on outcomes: grade level, race/ethnicity, FRM eligibility, current year attendance rate, MAP math percentile scores from September of both the current and prior year, MAP reading percentile scores from September of both the current and prior year, ELA course level (advanced or grade-level) and math ability level.

In this study, a grade level of 0 indicated a seventh-grade student and a grade level of 1 indicated an eighth-grade student. In addition, in this study, students were coded as 1 for low math ability, 2 for average math ability and 3 for high math ability. FRM eligibility of 0 indicated that a student was not eligible for free or reduced meals and FRM eligibility of 1 indicated that a student was eligible to receive free or reduced meals.

Research Design

A quasi-experimental research design was employed in this study to determine whether participation in the treatment group, advanced math course enrollment, improved students' education outcomes. Selection bias was a consideration in the research design. Enrollment in the treatment course is selective and requires students to meet both meritocratic and non-meritocratic criteria. Students who opt in to the treatment course must also be motivated and must self-advocate, or have parents or guardians advocating on their behalf. This selection bias might prove problematic if student achievement outcomes are attributed to participation in the treatment group when in fact they are the result of other factors, such as student motivation (Coca, Johnson, Kelley-Kemple, Roderick, Moeller, Williams & Moragne, 2012).

To account for potential selection bias, this study used propensity score matching to establish appropriate comparison groups of students in order to estimate the effect of participation in the treatment group. Prior studies have shown that propensity score matching is an effective method of reducing the impact of selection bias in estimating treatment effects (Voight & Velez, 2018). Studies also show that including a pretest measure of the outcome as an observed characteristic is among the most effective ways to address selection bias (Steiner, Cook, Shadish & Clark, 2010). This study included a pretest measure of the student achievement outcome in all models, which is explained in more depth in the analytic plan section below.

Prior research has also shown that propensity score matching is an effective method to estimate causal effects with observational data (Guo & Fraser, 2015; Rosenbaum & Rubin, 1983; Schneider, Carnoy, Kilpatrick, Schmidt & Shavelson, 2007;

Voight & Velez, 2018). The propensity score matching process includes several steps. First, for each research question, a logistic regression model was constructed to identify the predicted probability that students would enroll in an advanced math course. The logistic regression model included as predictors race, socioeconomic status, attendance and prior academic achievement, as these variables all have a significant impact on student achievement and might therefore distinguish students who are enrolled in advanced-level courses from those who are not (Domina & Saldana, 2012; Oakes, 2005; Steele, 2011; Voight & Velez, 2018).

The logistic regression equation predicts for each participant the probability of that student being enrolled in the advanced-level math course. The predicted probability is called a propensity score, and it takes continuous values between zero and one (Domina & Saldana, 2012). It is important to note that propensity score matching adjusts only for observed characteristics and may not account for alternative explanations. For this reason, this study included a relatively wide range of variables in order to preclude relevant factors from being eliminated (Schneider et al., 2007).

The second step of the propensity score matching process involves matching each participant to another in the opposite treatment group. Following the logistic regression analysis, treatment participants were matched with non-treatment participants with the most similar propensity score. The non-treatment participant group functions as a control group by providing an estimate of the estimated achievement outcomes if the student in the treatment group had in fact been enrolled in the non-treatment group (Domina & Saldana, 2012; Rosenbaum & Rubin, 1985).

Following the matching process, a post-matching balance test was conducted. The test assesses measured covariate balance to determine how closely covariates were matched during the propensity score process. Balancing measured covariates reduces overt bias in the estimation of the treatment effect and is a critical part of the propensity score matching process (Rosenbaum, 2002; Rubin & Thomas, 1996).

Analytic Plan

For each sample student, a propensity score was estimated to determine the student's likelihood of participating in an advanced math course in grade seven or eight. Propensity scores were estimated using the following logistic regression model equation:

$$ln\left(\frac{P_i}{1 - P_i}\right) = \beta_0 + \sum_{k=1}^{4} \beta_k Baseline_i + \sum_{l=5}^{10} \beta_l Demographics_i$$

In this model, P is the probability of student i participating in the treatment group and is a linear function calculated based on student baseline and demographic data. Baseline data include MAP math percentile scores from September of both the current and prior year, MAP reading percentile scores from September of both the current and prior year. Demographic data include grade level, race/ethnicity, FRM eligibility, current year attendance rate, ELA course level and math ability level. Prior research supports the inclusion of these covariates in this model, since race, socioeconomic status, attendance and prior academic achievement all have a significant impact on current student achievement (Domina & Saldana, 2012; Oakes, 2005; Steele, 2011; Voight & Velez, 2018).

Stata 14 was used to estimate propensity scores and to conduct the matching process. In order to ensure stronger measured covariate balance, caliper adjustments were used for some treatment variables. Caliper adjustments require that the difference

in propensity score between a treatment and non-treatment participant must be no greater than .25 standard deviations of all estimated propensity scores—called a "caliper" (Voight & Velez, 2018). For this study, matching with replacement was also used for some treatment variables. Matching with replacement allows each unit to be used as a match more than once. Prior studies have shown that matching with replacement produces higher quality matches, because it increases the set of possible matches (Abadie & Imbens, 2006). Students who were not matched within the caliper were not included in the final estimation of treatment effects.

The logistic regression model indicates which of the study variables have a statistically significant relationship to the propensity score outcome variable. In addition, the R squared value is reported in the findings below, since this figure estimates the amount that the combined predictor variables explain the variance in the likelihood of enrolling in an advanced math course. In order to assess the degree to which the matched groups are similar on the observed covariates, based on the observed characteristics, a series of post-match balance tests were conducted (Rosenbaum & Rubin, 1985). The effect of participating in the treatment course was measured by estimating the average treatment effects (ATE). The ATE measures the difference in mean outcomes between similar students who are enrolled in advanced-level courses and students who are not enrolled in the treatment group.

For each study research question, a series of parallel propensity score matching models were estimated. In each, propensity scores were calculated based on baseline and demographic data. Baseline data included MAP math percentile scores from September of both the current and prior year, MAP reading percentile scores from September of both

the current and prior year. Demographic data included grade level, race/ethnicity, FRM eligibility, current year attendance rate, ELA course level and math ability level. Once propensity scores were determined and students were matched, the effect of participation in the treatment course was calculated to compare a treated student's achievement outcomes with his or her non-treated student match's achievement outcomes. Student achievement outcomes for this research question were measured by student percentile scores in mathematics in May of the current year.

Research question 1 compared the achievement outcomes of students in advanced math classes to the achievement outcomes of students with comparable ability levels and background characteristics in grade-level math classes. To answer this question, three separate propensity score analyses compared the following: all students in advanced math classes with all students in grade-level math classes; all average-ability math students in advanced math classes with all average-ability students in grade-level math classes; and, all high-ability math students enrolled in advanced math classes with all high-ability students enrolled in grade-level classes.

Research question 2 compared the achievement outcomes of black students enrolled in advanced math classes to the achievement outcomes of black students with comparable ability levels and background characteristics enrolled in grade-level math classes. To answer this question, three separate propensity score analyses compared the following: all black students in advanced math classes with all black students in grade-level math classes; all average-ability black students in advanced math classes with all average-ability black students in grade-level math classes; and, all high-ability black

students enrolled in advanced math classes with all high-ability black students enrolled in grade-level classes.

Research question 3 compared the achievement outcomes of white students enrolled in advanced math classes to the achievement outcomes of white students with comparable ability levels and background characteristics enrolled in grade-level math classes. To answer this question, three separate propensity score analyses compared the following: all white students in advanced math classes with all white students in grade-level math classes; all average-ability white students in advanced math classes with all average-ability white students in grade-level math classes; and, all high-ability white students enrolled in advanced math classes with all high-ability white students enrolled in grade-level classes.

Summary

This chapter outlined how the study was conducted. A detailed explanation was provided of the study sample, the data collected and how propensity score matching was used for data analysis.

CHAPTER IV

RESULTS

This research explored the relationship between math tracking, race and student achievement outcomes. This chapter provides descriptive statistics in addition to findings that align with the study's three research questions: 1) How do the achievement outcomes of students enrolled in advanced math classes compare to the achievement outcomes of students with comparable ability levels and background characteristics who are enrolled in grade-level math classes? 2) How do the achievement outcomes of black students who are enrolled in advanced math classes compare to the achievement outcomes of black students with comparable ability levels and background characteristics who are enrolled in grade-level classes? and 3) How do the achievement outcomes of white students who are enrolled in advanced math classes compare to the achievement outcomes of white students with comparable ability levels and background characteristics who are enrolled in grade-level classes?

Descriptive Statistics

  Study sample demographics. This study included data from 1,510 students who were enrolled in grade 7 or 8 in Shaker Heights during the 2016-2017 or 2017-2018 school year (see Table I). The majority of the study sample was male (50.79 percent) and a plurality was black (43.58 percent). The study sample was evenly split between students enrolled in grade seven and students enrolled in grade eight. 12.52 percent, or 189, of students in the study sample had an Individualized Education Plan (IEP). In addition, 433 students (28.68 percent) in the sample qualified for free or reduced meals (FRM).

Table I

Frequency and percentages of sample data by gender, ethnicity, grade and demographic Information (n = 1,510)

Demographic Information	Frequency	Percentage
Gender		
Male	767	50.79
Female	743	49.21
Ethnicity		
White	610	40.40
Black	658	43.58
Other	197	13.05
Grade		
Grade 7	755	50.00
Grade 8	755	50.00
Demographic information		
IEP	189	12.52
FRM eligible	433	28.68

Table II displays central tendency data of current and prior year MAP math and reading percentile scores, as well as current and prior year attendance rate.

Table II

Central tendency data of continuous variables current (CY) and prior year (PY) MAP math and reading percentile scores and CY attendance (att.) rate

	Obs.	Min.	Max.	Mean	Median	S.D.
CY Sept. MAP math percentile scores	1,112	1	99	65.15	72	26.21
CY Sept. MAP reading percentile scores	1,112	1	99	68.18	74	24.89
CY att. rate	1,127	.46	1	.96	.97	.04
PY Sept. MAP math percentile scores	1,052	1	99	66.04	72	24.19
PY Sept. MAP reading percentile scores	1,049	1	99	69.45	76	23.52

In September of the current year, the mean percentile score on the MAP mathematics test was 65.15 and the mean percentile score on the MAP reading test during this time was 68.18. The median percentile score in September of the current year on the MAP mathematics test was 72 and the median percentile score on the MAP reading test during this time was 74. The standard deviation of the MAP math percentile score in September of the current year was 26.21 and, on the MAP reading test, the

standard deviation was 24.89 in September of the current year. In the current year, the mean attendance rate was .96, the median attendance rate was .97 and the standard deviation was .04.

In September of the prior year, the mean percentile score on the MAP mathematics test was 66.04 and the mean percentile score on the MAP reading test during this time was 69.45. The median percentile score in September of the prior year on the MAP mathematics test was 72 and the median percentile score on the MAP reading test during this time was 76. The standard deviation of the MAP math percentile score in September of the prior year was 24.19 and, on the MAP reading test, the standard deviation was 23.52 in September of the prior year.

Table III shows these data broken down by black and white student subgroups.

Table III

Central tendency data of black and white student CY and PY MAP math and reading percentile scores and CY att. rate

	Black student obs.	Black student Mean	Black student Median	Black S.D.	White student obs.	White student Mean	White student Median	White S.D.
CY Sept. MAP math percentile scores	449	47.23	47	24.33	498	79.85	84	17.03
CY Sept. MAP reading percentile scores	449	52.76	55	24.50	498	80.91	86	17.05
CY att. rate	460	.96	.97	.04	496	.96	.97	.04
PY Sept. MAP math percentile scores	429	49.58	50	22.64	483	78.62	83	16.81
PY Sept. MAP reading percentile scores	425	55.22	57	23.39	484	80.92	86	16.41

Table III illustrates discrepancies between black and white student achievement outcomes. For example, in September of the current year, the mean math percentile score for black students was 47.23, as compared to a mean score of 79.85 for white students. A similar gap is evident in current year MAP reading percentile scores. For example, in

112

September of the current year, the mean black student score was 52.76, as compared to a mean score of 80.91 for white students. The number of black student observations is consistently slightly less than the number of white student observations. These data are slightly different from the overall percentage of black students as listed in Table I. This discrepancy may be accounted for by the fact that a higher number of black students were not included in this data set, because they may have been enrolled in classes, such as special education classes, in which they did not take the MAP assessment.

One area in which black and white student data are similar is attendance rate data. In the current year, both black students and white students had a mean attendance rate of .96. These data show that black and white students have similar attendance patterns.

The data in Table III underscore the difficulty in a simple comparison of black and white student achievement outcomes. On average, white students achieve percentile scores that are nearly 40 percent higher than their black counterparts. Consequently, simply comparing black student achievement outcomes with white student achievement outcomes would not fully account for the discrepancies that exist in black-white student achievement from the outset, since the two subgroups of students differ drastically in their initial achievement metrics.

Data in Table IV show current and prior year student enrollment in math and ELA courses, divided by ethnicity category. Like the data in Table III, these data illustrate discrepancies that exist between black and white student subgroups.

Table IV

Black and white student enrollment in CY and PY math and English Language Arts (ELA) courses and prior math ability levels

Demographic information	Total student frequency	Black student frequency	Black student percentage	White student frequency	White student percentage
CY math enrollment					
Grade level	491	339	69.04	83	16.90
Advanced	641	122	19.03	416	64.90
CY ELA enrollment					
Grade level	472	321	68.01	82	17.37
Advanced	662	140	21.15	418	63.14
Prior math ability level					
Low	81	71	87.65	5	6.17
Average	519	305	58.77	157	30.25
High	540	88	16.30	340	62.96

According to these data, in the current year, 339 black students were enrolled in grade-level mathematics in grade seven or eight. These 339 students comprised 69.04 percent of all students enrolled in grade-level mathematics. By contrast, only 83 white students were enrolled in a grade-level mathematics course in the current year. These 83 students comprised 16.90 percent of all students enrolled in grade-level mathematics courses. Figures for enrollment in advanced-level courses are nearly the exact opposite. In the current year, 122 black students were enrolled in advanced-level mathematics courses in grades seven or eight, as compared to 416 white students who were enrolled in advanced-level mathematics. These numbers represent 19.03 percent and 64.90 percent

of all students enrolled in advanced-level mathematics, respectively. A similar pattern exists for ELA course enrollment.

Differences in prior ability levels reflect the disparities in course enrollment. According to the figures shown in Table IV, 71 black students have low math ability, 305 have average ability and 88 have high ability. Put another way, 65.73% of black students have average math ability and 18.97% of black students have high math ability. By contrast, only 5 white students have low math ability and 157 have average ability. The remainder of white students, 340, are considered to be in the high-ability category. Based on these numbers, 67.73% of white students are high-ability students in mathematics and 31.27% of white students are average-ability math students. Only one percent of white students are low ability, as compared to 15.30% of black students. These significant discrepancies in these two subgroups underscore again the limitations in simply comparing black and white student achievement outcomes without taking into account baseline differences between the two subgroups.

Data in Table V show current year student enrollment in math course levels, divided by ability group. Like the previous tables, this table shows discrepancies between black and white student subgroups. For example, only 56.32 percent of high-ability black students are enrolled in an advanced-level math class, as compared to 96.17 percent of high-ability white students.

Table V

Black and white student enrollment in CY math course level, by ability level

	Black student enrollment				White student enrollment			
Prior math ability level	Grade level freq.	Grade level pct.	Adv. freq.	Adv. pct.	Grade level freq.	Grade level pct.	Adv. freq.	Adv. pct.
Low	69	100.00	0	0	4	100.00	0	0
Avg	232	76.07	73	23.93	66	42.31	90	57.69
High	38	43.68	49	56.32	13	3.83	326	96.17

Outcome Variable

In this study, the outcome variable was the current year May MAP math percentile score. Table VI provides central tendency data of the outcome variable. Table VI shows the range, mean, median and standard deviation of the outcome variable for all students and broken down by black and white student subgroups.

Table VI

Central tendency data of Outcome Variable divided by student subgroups

	Obs.	Min.	Max.	Mean	Median	S.D.
Overall CY May MAP math percentile scores	1,124	1	99	67.40	75	26.75
Black student May MAP math percentile scores	454	1	99	49.34	48	24.81
White student May MAP math percentile scores	500	1	99	82.95	89	16.25

116

According to the figures in Table VI, 1,124 students received percentile scores on the May MAP math assessment in the current year. These students scored a mean percentile score of 67.40, with a median score of 75 and a standard deviation of 26.75. Table VI also clearly shows that black students performed worse than their white counterparts. The mean percentile score for black students on the May MAP math assessment in the current year was 49.34, as compared to a mean percentile score of 82.95 for white students. The standard deviation for black student percentile scores on the May MAP math assessment in the current year was 24.81, as compared to a standard deviation of 16.25 for white students. This number indicates that black student percentile scores are spread out over a wider range of values away from the mean than white students.

Models

This study estimated a series of propensity score matching models in order to answer the research questions. Several treatment variables were created in order to establish appropriate comparison groups of students and to estimate the effect of participation in the treatment group. The treatment group for model 1 is comprised of all students who were enrolled in an advanced-level math class during the current academic year. The control group for model 1 is comprised of students who were enrolled in grade-level classes. The treatment group for model 2 includes students who were enrolled in an advanced-level math course and who have an average prior math ability level. The control group for model 2 is comprised of average-ability students who took a grade-level math course. The treatment group for model 3 is comprised of high-ability

students who were enrolled in an advanced-level math course. Students of high ability who were enrolled in a grade-level class comprise the control group for model 3.

Subsequent treatment variables break students up by racial subgroup. Treatment group 4 includes black students enrolled in an advanced-level math course. The control group for model 4 includes black students who were enrolled in a grade-level math course. Treatment group 5 includes average-ability black students who were enrolled in an advanced-level math course. The control group for model 5 is comprised of average-ability black students who took a grade-level math course. Treatment group 6 is comprised of high-ability black students were enrolled in an advanced-level math course. Black students of high ability who were enrolled in a grade-level class comprise the control group for model 6.

Treatment group 7 includes white students enrolled in an advanced-level math course. The control group for model 7 includes white students enrolled in a grade-level math course. Treatment group 8 includes average-ability white students enrolled in an advanced-level math course. The control group for model 8 is comprised of average-ability white students who took a grade-level math course. Finally, treatment group 9 is comprised of high-ability white students were enrolled in an advanced-level math course. White students of high ability who were enrolled in a grade-level class comprise the control group for model 9.

Table VII shows the frequencies for the control and treatment groups for the study treatment variables.

Table VII

Treatment and control group frequencies, by model

Model	Sample	Control group frequency	Treatment group frequency
1	All students	491	641
2	All average-ability students	337	181
3	All high-ability students	29	445
4	Black students	339	122
5	Average-ability black students	232	73
6	High-ability black students	38	49
7	White students	83	416
8	Average-ability white students	66	90
9	High-ability white students	13	326

Note: Control group = grade-level math course; Treatment group = advanced-level math course

The data in Table VII show the wide variation in treatment and control group size. For example, in model 1, the treatment group is composed of 641 students and the control group is composed of 491 students. Treatment group 9, however, which is composed of high-ability white students enrolled in advanced-level courses, includes 326 students as compared to only 13 students in the control group. Given the discrepancies in control group and treatment group size, and the smaller sample size of some of the groups, a propensity score matching model was used to compare the achievement outcomes of students of comparable ability.

Descriptive Comparisons of Treatment v. Non-Treatment Students

   A descriptive comparison between treatment and non-treatment students was conducted prior to conducting propensity score matching. The descriptive comparison included a comparison of treatment and non-treatment students in terms of the demographic composition of the two groups as well as their mean Map math percentile scores and prior math ability levels. The comparison is included in Table VIII.

Table VIII

Descriptive comparison of treatment v. non-treatment students by demographic data, pre-matching

Model	Grade level	Black	White	Math ability level	CY att. rate	FRM
Model 1						
Control	.5397	.7182	.1758	1.9959	.9519	.4678
Treatment	.4789	.1937	.6603	2.7176	.9642	.0922
Model 2						
Control	.5490	.6884	.1958	2	.9549	.4451
Treatment	.3867	.4056	.5000	2	.9638	.1713
Model 3						
Control	.4483	.4138	.4830	3	.9503	.2069
Treatment	.5303	.1103	.7241	3	.9646	.0608
Model 4						
Control	.5398	1	0	1.9086	.9540	.5629
Treatment	.4262	1	0	2.4016	.9717	.3279
Model 5						
Control	.5603	1	0	2	.9568	.5474
Treatment	.3562	1	0	2	.9707	.3425
Model 6						
Control	.5000	1	0	3	.9451	.4848
Treatment	.5306	1	0	3	.9733	.3061
Model 7						
Control	.5904	0	1	2.1084	.9444	.0732
Treatment	.5000	0	1	2.7837	.9616	.0265
Model 8						
Control	.5303	0	1	2	.9490	.0758
Treatment	.4222	0	1	2	.9601	.0444
Model 9						
Control	.9231	0	1	3	.9157	0
Treatment	.5215	0	1	3	.9620	.02154

The data in Table VIII provide a descriptive comparison between treatment and control groups for the various models. The table highlights many of the differences between the treatment and control groups. For example, treatment group 1 represents all students who were enrolled in advanced-level mathematics courses. The composition of the treatment and control groups for model 1 is distinctly different in terms of black and white student enrollment. 71.82 percent of the control group in model 1 is comprised of black students as compared to 19.36 percent of the treatment group. This is nearly the opposite for white students, who comprise 17.58 percent of the control group and 66.03 percent of the treatment group for model 1.

This type of disparity, in terms of black and white student make-up of the treatment and control groups, exists for many of the study models. For example, treatment group 3 represents students of high math ability who were enrolled in advanced courses. In this model, black students comprise 11.03 percent of the treatment group, while white students comprise 72.41 percent of the treatment group. Control group 3 is comprised of high ability math students who are enrolled in grade-level math classes. The composition of the control group 3 is more evenly distributed between black and white students, since 41.38 percent of the control group is made up of black students and 48.28 percent of white students.

Several of the treatment variables include only one racial subgroup. For example, treatment group 8 is comprised of average-ability white students enrolled in an advanced-level math course. The control group for treatment variable 8 is made up of average-ability white students enrolled in a grade-level math course. Similarly, treatment group 4 is made up of black students enrolled in an advanced-level math course and the control

group for treatment variable 4 is comprised of black students enrolled in a grade-level math course.

Treatment and non-treatment students also differ by prior math ability level. For example, in model 4, the control group had a mean ability level of 1.9086 and the treatment group had a mean ability level of 2.4016. According to the data in Table VIII, the greatest disparity in ability level exists in model 1. The mean math ability level in control group 1 is 1.9959, as compared to a mean math ability level of 2.7176 in the treatment group. This is not surprising, since one would expect that students who are enrolled in grade-level classes would have a lower math ability level, on average, than students who are enrolled in advanced-level math courses. Nevertheless, as noted previously, disparities in terms of the background characteristics of treatment and non-treatment students make a simple descriptive comparison between groups insufficient to fully answer this study's research questions.

Treatment and control groups also differ by student socio-economic status, as measured by eligibility for FRM. For example, in model 1, only 9.22 percent of the treatment group is comprised of students who are eligible for FRM, as compared to 46.78 percent of the control group. Treatment group 5 includes average-ability black students enrolled in advanced-level courses. 54.74 percent of the control group is eligible for FRM, as compared to 34.25 of the treatment group.

Several factors are more similar across treatment and control groups. For example, students in treatment and control groups across treatment variables have similar attendance rates, though students enrolled in advanced-level math classes have consistently higher attendance rates across all models. Attendance rates were calculated

by dividing total days enrolled by total days present. In model 8, the mean attendance rate for the control group was .9490 and the mean attendance rate for the treatment group was .9601. In treatment variable 1, the mean attendance rate for the control group was .9519 and for the treatment group was .9642.

The mean grade level varies across treatment and control groups. In treatment variable 2 for example, the mean grade level for the control group was .5490, which means there were more eighth-grade students than seventh-grade students. The mean grade level for the treatment group, however, was .3867, which indicates that there were more seventh-grade students in the treatment group than eighth-grade students.

As a whole, the descriptive comparisons that are displayed in Table VIII indicate that there are substantial differences between the treatment and control groups, particularly in terms of black and white student composition. Consequently, it would be difficult to easily compare these two groups through an independent samples t test. The propensity score matching model is more suited to comparing achievement outcomes between the treatment and control groups, because it allows for students of similar ability and demographic background to be compared against one another.

Table IX shows a comparison of baseline math assessment scores of mean student CY and PY September MAP math assessment percentile scores, as well as mean percentile scores of the outcome variable, the CY May MAP math assessment, pre-matching. Similar to Table VIII, these data reflect disparities between the treatment and control groups across covariates.

Table IX

Descriptive comparison of treatment v. non-treatment students by baseline and outcome MAP math percentile scores, pre-matching

| Model | Baseline Data | | Outcome Variable |
	PY Sept. MAP math percentile score	CY Sept. MAP math percentile score	CY May MAP math percentile score
Model 1			
Control	45.09	42.42	43.76
Treatment	81.02	82.38	85.28
Model 2			
Control	49.62	47.40	49.01
Treatment	66.38	68.82	72.92
Model 3			
Control	77.62	61.90	64.07
Treatment	87.32	87.94	90.35
Model 4			
Control	41.27	38.22	39.74
Treatment	72.13	72.83	76.19
Model 5			
Control	47.07	43.61	45.57
Treatment	65.59	65.77	69.57
Model 6			
Control	65.75	39.96	39.84
Treatment	82.73	84.04	85.92
Model 7			
Control	56.67	56.78	60.38
Treatment	83.10	84.65	87.43
Model 8			
Control	55.52	56.61	60.26
Treatment	67.26	71.21	75.44
Model 9			
Control	82.22	65.92	73.33
Treatment	87.68	88.41	90.74

The data in this table underscore again the disparities that exist between the treatment and control groups across models. For example, in model 1, the mean PY September MAP math assessment percentile score was 45.09 for students in the control group and 81.02 in the treatment group. The mean CY September MAP math assessment percentile score was 42.42 for students in the control group and 82.38 in the treatment group. In this model, the mean CY May MAP math percentile score for the control group was 43.76, which represents an increase of 1.34 from the CY September MAP math mean percentile score. The mean percentile score for the treatment group was 85.28, which represents an increase of 2.90 from the CY September MAP math mean percentile score. On its face, therefore, it appears that students in the treatment group performed better over time than students in the control group. Given the disparities in both the demographic and baseline composition of these two groups, however, a descriptive comparison is insufficient to fully answer the research questions and to determine how students of comparable ability fared over time.

Estimates of Treatment Effects

The findings from the nine propensity score matching models regarding the effect of enrolling advanced-level math courses are presented below in conjunction with the appropriate research question.

Research Question 1

How do the achievement outcomes of students enrolled in advanced math classes compare to the achievement outcomes of students with comparable ability levels and background characteristics who are enrolled in grade-level math classes?

126

Covariate balance between matched pairs. Research question 1 included

models 1 through 3. Initial matching procedures for model 1 resulted in a total of 1,005

students matched for the test of average treatment effects of advance-level math

enrollment on May MAP math percentile scores. 595 students were enrolled in

advanced-level math and 410 students were in the control group. For model 1, the

modeled covariates include: black, other ethnicity, grade level, attendance rate, FRM,

MAP math percentile scores from September of both the current and prior year, MAP

reading percentile scores from September of both the current and prior year, ELA course

level and math ability level.

Checking balance in covariates across treatment and control groups is important,

because propensity score matching assumes that cases with the same propensity score

have the same distributions for observable and non-observable characteristics (Stone &

Tang, 2013). In this analysis, results of tests of covariate balance generally indicated

good balance between groups in both samples for model 1. These results are supported

by variance ratio tests and standardized difference tests and are shown in Table X. All

variance ratios for all covariates fall between .5 and 2.0 after being matched. For

example, the matched variance ratio of the current year September MAP math percentile

score is 1.44 and the matched variance ratio for math ability level is .67.

The absolute value of all standardized differences for most matched covariates are

less than .2. For example, the matched standardized difference for ability level is .14 and

the matched standardized difference for prior year September MAP math percentile

scores is .07. As Table X shows, however, the standardized difference for several

matched covariates did fall out of the .2 range. For example, the matched standardized

difference for English course level is .36 and the matched standardized difference for the prior year September MAP reading percentile score is .41. In these cases, however, the matched standardized difference was substantially lower than the raw standardized difference, which indicates that the covariates were more comparable following the test application than prior to the test application.

The treatment group for model 2 is comprised of all average-ability students who were enrolled in an advanced-level math class during the current academic year and the control group for model 2 is comprised of all average-ability students enrolled in grade-level classes. For model 2, the modeled covariates include: black, other ethnicity, grade level, attendance rate, FRM, MAP math percentile scores from September of both the current and prior year, MAP reading percentile scores from September of both the current and prior year, ELA course level and math ability level. Initial matching procedures for model 2 resulted in a total of 499 students matched for the test of average treatment effects for May MAP math percentile scores. 177 students were in the treatment group and 322 students were in the control group.

In the analysis of model 2, results of tests of covariate balance generally indicated good balance between groups in both samples. As shown in Table X, variance ratio tests and standardized difference tests supported these results. All variance ratios for matched covariates fall between .5 and 2.0. The absolute value of all standardized differences for most matched variables are also less than .2, though, similar to model 1, the standardized difference for several matched covariates did fall slightly out of the .2 range. For example, the matched standardized difference for attendance rate is .30 and the matched standardized difference for FRM status was -.28. Again, the matched standardized

difference was substantially lower than the raw standardized difference, which indicates that the covariates were more comparable following the test application than prior to the test application.

Table X

Covariate balance summary for models 1-2

	Model 1				Model 2			
	Std. Differences		Variance ratio		Std. Differences		Variance ratio	
Covariate	Raw	Match	Raw	Match	Raw	Match	Raw	Match
Black	-1.20	.07	.76	1.03	-.56	-.04	1.11	1.01
Other (ethnicity)	.14	-.12	1.35	.79	-.10	.10	.77	1.24
Grade level	-.15	-.19	1.00	1.02	-.32	-.16	.95	.97
Att. Rate	.29	-.10	.69	1.13	.21	.30	.93	.92
FRM	-.93	-.06	.34	.92	-.62	-.28	.57	.75
CY MAP math percentile score Sept.	2.24	.12	.46	1.44	1.30	-.12	.68	1.78
CY MAP read percentile score Sept.	1.45	.35	.44	.67	.69	.01	.66	.74
PY MAP math percentile score Sept.	2.14	.07	.42	.89	1.20	-.09	.53	1.21
PY MAP read percentile score Sept.	1.51	.41	.46	.74	.77	.04	.74	1.06
CY ELA course level	1.70	.36	.70	.86	1.01	.13	1.13	1.02
Math ability	1.86	.14	1.07	.66				

The treatment group for model 3 includes all high-ability students who were enrolled in an advanced-level math class during the current academic year and the control group includes all high-ability students who were enrolled in grade-level classes. For model 3, the modeled covariates include: black, other ethnicity, grade level, attendance rate, FRM, MAP math percentile scores from September of both the current and prior year, MAP reading percentile scores from September of both the current and prior year, ELA course level and math ability level. Initial matching procedures for model 3 resulted in a total of 410 students matched for the test of average treatment effects for May MAP math percentile scores. 394 students were in the treatment group and 16 students were in the control group. In this case, the small sample size of the control group is not surprising, given that one would expect most high-ability students, who score, on average, above the 75th percentile score, to be enrolled in advanced-level math courses.

The results of tests of covariate balance for model 3 showed weak and inadequate balance between matched covariates. For example, the matched standardized difference of pretest math data fell well above .2. The serious imbalance in matched covariates, particularly in terms of pretest math data, renders the estimation of treatments effects deeply susceptible to bias and problematic when considering validity of the findings. As a result, though the PS match for model 3 generated statistically significant results, the covariate matched statistics are not reported in these findings.

Analysis of excluded cases. Due to failure to match with an opposite condition case within the specified parameter, 14 observations were discarded from tests of the treatment effect on student achievement outcomes. These excluded students had

significantly lower attendance rates, MAP math and reading percentile scores and English course levels and significantly higher rates of FRM eligibility ($p < .05$, based on independent samples t tests). Further, students with Individualized Education Plans (IEPs) were discarded from the analysis, because of small sample size and because these students were significantly different from the matched cases, in the same ways described above.

Caliper adjustments and matching with replacement. For model 1, treatment effects were estimated with no caliper. For models 2 and 3, treatment effects were estimated using matching with narrow calipers, calculated at .25 propensity-score standard deviation units. Caliper adjustments for models 2 and 3 ensured a stronger post-match covariate balance (Caliendo & Kopeinig, 2008).

For models 1 and 3, matching was conducted with replacement, allowing an untreated student to be matched up to three times. For model 2, matching was conducted without replacement. Matching with replacement was used in order to ensure a stronger post-match covariate balance (Caliendo & Kopeinig, 2008).

Logistic regression results. For each research question, a logistic regression model was constructed to identify the predicted probability that individuals with certain characteristics would be assigned to the treatment group. The regression model included the study variables, because prior research demonstrates that race, socioeconomic status, attendance and prior academic achievement all have a significant impact on student achievement and might therefore distinguish students who are enrolled in advanced-level courses from those who are not (Domina & Saldana, 2012; Oakes, 2005; Steele, 2011; Voight & Velez, 2018).

The logistic regression analysis indicates which of the study variables have a statistically significant relationship to the propensity score treatment variable. Results for the logistic regression analyses conducted for models 1 through 3 are listed below, in Table XI. Current and prior year September MAP math percentile scores are shown to consistently have a statistically significant relationship with the likelihood of being enrolled in the treatment group, across all three models. Grade level also has a statistically significant relationship with the likelihood of being in the treatment group, across all three models. For models 1 and 2, current year attendance rate, FRM eligibility and ELA course level have a statistically significant relationship with the likelihood of being in the treatment group.

For model 1, the R squared value is .64, which indicates that the combined predictor variables explain approximately 64 percent of the variance in the likelihood of enrolling in an advanced math course. For model 2, the R squared value is .39, which indicates that the combined predictor variables explain approximately 39 percent of the variance in the likelihood of enrolling in an advanced math course for students of average ability. For model 3, the R squared value is .21, which indicates that the combined predictor variables explain approximately 21 percent of the variance in the likelihood of enrolling in an advanced math course for high-ability math students.

Table XI

Logistic Regression Results for Models 1-3

Predictor	Model 1		Model 2		Model 3	
	Coef. (β)	Std. Error	Coef. (β)	Std. Error	Coef. (β)	Std. Error
Black	-.36	.28	-.33	.30	-.04	.81
Other (ethnicity)	-.29	.39	-.42	.44	-.22	.70
CY grade level	-1.34***	.25	-1.17***	.55	1.34*	059
CY att. rate	9.51***	2.99	7.78*	3.77	7.84	6.84
FRM	-.66*	.30	-.66*	.32	.03	.95
CY MAP math percentile score Sept.	.06***	.01	.06***	.01	.06**	.02
CY MAP read percentile score Sept.	-.01	.01	-.01	.01	-.02	.03
PY MAP math percentile score Sept.	.04***	.01	.04***	.01	.80**	.04
PY MAP read percentile score Sept.	-.00	.01	.00	.01	-.00	.03
ELA course level	1.32***	.28	1.22***	.30	.48	.82
Math ability level	1.21**	.39				
R squared	.64		.39		.21	

***$p < .001$, **$p < .01$, *$p < .05$

Average treatment effects. Average treatment effect results are presented in terms of unadjusted measurement units and in terms of effect size (ES). The ES is measured by Cohen's *d*, which is the difference between the mean outcome for the

treatment group and the mean outcome for the control group divided by the standard deviation. A *d* of less than or equal to .2 indicates a small ES, a *d* between .2 and .8 indicates a medium ES and a *d* of greater than or equal to .8 indicates a large ES (Warner, 2013).

Data from the PS analysis show that the average treatment effect was statistically significant across models 1 and 2. As shown in Table XII, these data indicate that participation in an advanced-level math course had a statistically significant impact on the student achievement outcome variable for all students and average-ability students. For example, treatment group 1, which is comprised of all students enrolled in an advanced-level math course, is associated with a percentile increase of 11.86 points ($p <$ .01). For average-ability students, participating in an advanced-level math course is associated with an increase of 9.40 percentile points ($p <$ .01). The *d* for treatment variables 1 and 2 shows a moderate ES.

Table XII

Estimated effects of advanced-level math enrollment on math achievement outcomes for models 1-2

Model	*n*	Coef.	*SE*	Effect Size (d)
Model 1	1,005	11.86***	2.40	.46
Model 2	499	9.40**	3.53	.47

***$p <$ .001, **$p <$ .01, * $p <$ .05

Research Question 2

How do the achievement outcomes of black students who are enrolled in advanced math classes compare to the achievement outcomes of black students with

comparable ability levels and background characteristics who are enrolled in grade-level classes?

Covariate balance between matched pairs. Research question 2 was investigated through a PS matching model that included models 4 through 6. The treatment group for model 4 is comprised of all black students who were enrolled in an advanced-level math class during the current academic year. The control group for model 4 is comprised of all black students who were enrolled in grade-level classes. Initial matching procedures for model 4 resulted in a total of 406 students matched for the test of average treatment effects for May MAP math percentile scores. 116 students were enrolled in advanced-level math and 290 students were in the control group.

For model 4, the modeled covariates include: grade level, attendance rate, FRM status, MAP math percentile scores from September of both the current and prior year, MAP reading percentile scores from September of both the current and prior year and math ability level. In this analysis, results of tests of covariate balance generally indicated moderate balance between groups in both samples for model 4. These results are supported by variance ratio tests and standardized difference tests as shown in Table XIII. All variance ratios for all covariates fall between .5 and 2.0 after being matched. For example, the matched variance ratio of the current year September MAP math percentile score is 1.36 and the matched variance ratio for math ability level is .41.

The absolute value of standardized differences for most matched covariates are less than .2. For example, the matched standardized difference for current year September MAP reading percentile scores is .01 and the matched standardized difference current year September MAP math percentile scores is -.09. As Table XIII shows,

however, the standardized difference for several matched covariates did fall out of the .2 range. For example, the matched standardized difference for prior year MAP math percentile score is .25, the matched standardized difference for the prior year September MAP reading percentile score is .46 and the matched standardized difference for prior math ability is .43. In all of these cases, however, the matched standardized difference was substantially lower than the raw standardized difference, which indicates that the covariates were more comparable following the test application than prior to the test application.

The treatment group for model 5 is comprised of all average-ability black students who were enrolled in an advanced-level math class during the current academic year and the control group for model 5 is comprised of all average-ability black students enrolled in grade-level classes. Initial matching procedures for model 5 resulted in a total of 290 students matched for the test of average treatment effects for May MAP math percentile scores. 72 students were in the treatment group and 218 students were in the control group.

For model 5, the modeled covariates include: grade level, attendance rate, FRM status, MAP math percentile scores from September of both the current and prior year, MAP reading percentile scores from September of both the current and prior year and math ability level. In the analysis of model 5, results of tests of covariate balance generally indicated good balance between groups in both samples. These results are supported by variance ratio tests and standardized difference tests and are shown in Table XIII. All variance ratios for matched covariates fall between .5 and 2.0. The absolute value of standardized differences for most matched variables is also less than .2, except

for prior year MAP reading percentile score which is .35. Again, however, the matched standardized difference was substantially lower than the raw standardized difference, which indicates that the covariates were more comparable following the test application than prior to the test application.

Table XIII

Covariate balance summary for models 4-5

| | Treatment variable 4 | | | | Treatment variable 5 | | | |
| | Std. differences | | Variance ratio | | Std. differences | | Variance ratio | |
Covariate	Raw	Match	Raw	Match	Raw	Match	Raw	Match
Grade level	-.26	-.20	.97	.98	-.40	-.08	.92	1.00
Att. Rate	.48	.26	.37	.54	.39	.20	.46	.66
FRM	-.49	-.28	.92	.96	-.45	-.12	.91	.98
CY MAP math percentile score Sept.	1.90	.09	.72	1.36	1.30	.03	.84	1.25
CY MAP read percentile score Sept.	1.39	.01	.50	1.13	1.03	-.11	.55	1.23
PY MAP math percentile score Sept.	1.86	.25	.59	.85	1.03	-.05	.80	1.75
PY MAP read percentile score Sept.	1.49	.46	.36	.33	1.27	.35	.40	.44
Math ability level	1.23	.43	1.14	.41				

The treatment group for model 6 includes all high-ability black students enrolled in an advanced-level math class during the current academic year and the control group includes all high-ability black students who were enrolled in grade-level classes. For model 6, the modeled covariates include: grade level, attendance rate, FRM status, MAP math percentile scores from September of both the current and prior year, MAP reading percentile scores from September of both the current and prior year and math ability

level. Initial matching procedures for model 6 resulted in a total of 51 students matched

for the test of average treatment effects for May MAP math percentile scores. 44

students were in the treatment group and 7 students were in the control group. In this

case, the small sample size of both the control group and treatment group is due to the

small number of black students who are identified as high ability as well as the relatively

small number of black students enrolled in advanced-level math classes.

The results of tests of covariate balance showed weak balance between matched

covariates. In addition, the PS match for model 6 generated statistically insignificant

results. As a result, the covariate matched statistics are not reported in these findings.

Caliper adjustments and matching with replacement. For models 4 and 5,

treatment effects were estimated using matching with narrow calipers, calculated at .25

propensity-score standard deviation units. For model 6, treatment effects were estimated

using matching with no caliper. Caliper adjustments for models 4 and 5 ensured a

stronger post-match covariate balance. For models 4, 5 and 6, matching was conducted

with replacement, allowing an untreated student to be matched up to three times.

Matching with replacement was used in order to ensure a stronger post-match covariate

balance (Caliendo & Kopeinig, 2008).

Logistic regression results. Results for the logistic regression analyses

conducted for models 4 through 6 are listed below, in Table XIV. For models 4 and 5,

CY and PY September MAP math percentile scores have a statistically significant

relationship with the likelihood of being enrolled in the treatment group. Grade and ELA

course level also has a statistically significant relationship with the likelihood of being in

the treatment group, across models 4 and 5. For model 6, no predictor was found to have a statistically significant relationship with the outcome variable.

For model 4, the R squared value is .56, which indicates that the combined predictor variables explain approximately 56 percent of the variance in the likelihood of black students enrolling in an advanced math course. For model 5, the R squared value is .42, which indicates that the combined predictor variables explain approximately 42 percent of the variance in the likelihood of enrolling in an advanced math course for black students of average ability. For model 6, the R squared value is .55, which indicates that the combined predictor variables explain approximately 55 percent of the variance in the likelihood of enrolling in an advanced math course for high-ability black math students.

Table XIV

Logistic Regression Results for Models 4-6

Predictor	Treatment variable 4		Treatment variable 5		Treatment variable 6	
	Coef. (β)	Std. Error	Coef. (β)	Std. Error	Coef. (β)	Std. Error
CY grade level	-1.48***	.39	-1.27**	.40		
CY att. rate	6.25	6.43	8.43	6.93	-23.26	30.51
FRM	-.53	.36	-.46	.38	-1.15	1.30
CY MAP math percentile score Sept.	.05***	.01	.05**	.01	.10	.06
CY MAP read percentile score Sept.	-.00	.01	-.00	.01	.02	.05
PY MAP math percentile score Sept.	.04*	.02	.04*	.02	.15	.10
PY MAP read percentile score Sept.	.01	.01	.02	.01	-.10	.08
ELA course level	1.51***	.39	1.34**	.41	1.38	1.35
Math ability level	1.01	.64				
R squared	.56		.42		.55	

***$p < .001$, **$p < .01$, * $p < .05$

Average treatment effects. Data from the PS analysis show that the average treatment effect was statistically significant across two of the three treatment variables. As shown in Table XV, results indicate that participation in an advanced-level math course had a statistically significant impact on the student achievement outcome variable.

142

For example, treatment group 4, which comprises black students enrolled in an advanced-level math courses, is associated with a percentile score increase of 13.36 points ($p <$ .01). For average-ability black students, participating in an advanced-level math course is associated with an increase of 9.89 percentile points ($p < .01$). As previously mentioned, the test results for treatment variable 6 were not statistically significant. *Cohen's d* for treatment variables 4 and 5 shows a moderate ES.

Table XV

Estimated effects of treatment group participation on student achievement outcomes for models 4-6

Model	n	Coef.	*SE*	Effect Size *(d)*
Model 4	406	13.36***	3.42	.54
Model 5	290	9.89***	3.64	.50
Model 6	51	7.47	9.10	.26

***$p < .001$

Research Question 3

How do the achievement outcomes of white students who are enrolled in advanced math classes compare to the achievement outcomes of white students with comparable ability levels and background characteristics who are enrolled in grade-level classes?

Covariate balance between matched pairs. Research question 3 was investigated using nearest-neighbor and PS matching models that included models 7 through 9. Nearest-neighbor matching determines similarity between subjects by identifying the nearest match based on a weighted function of the covariates for each observation. Like PS matching, the nearest-neighbor matching model matches subjects

based on a single continuous covariate, which is the estimated treatment probability (Abadie, Dukker, Herr & Imbens, 2004).

The treatment group for model 7 is comprised of all white students enrolled in an advanced-level math class during the current academic year. The control group for model 7 is comprised of all white students enrolled in grade-level math classes. Nearest-neighbor matching was conducted to match treated students with control group students. Initial matching procedures for model 7 resulted in a total of 464 students matched for the test of average treatment effects for May MAP math percentile scores. 388 students were enrolled in advanced-level math and 76 students were in the control group.

Measured covariates for model 7 include: grade level, attendance rate, FRM status, MAP math percentile scores from September of both the current and prior year, MAP reading percentile scores from September of both the current and prior year and math ability level. In this analysis, the results of tests of covariate balance for model 7 showed weak balance between matched covariates. The serious imbalance in matched covariates, particularly in terms of pretest math baseline data, renders the estimation of treatments effects deeply susceptible to bias and problematic when considering validity of the findings. As a result, though the PS match for model 7 generated statistically significant results, the covariate matched statistics are not reported in these findings.

The treatment group for model 8 is comprised of all average-ability white students who were enrolled in an advanced-level math class during the current academic year and the control group for this model is comprised of all average-ability white students enrolled in grade-level classes. PS matching procedures for model 8 resulted in a total of 154 students matched for the test of average treatment effects for May MAP math

percentile scores. 89 students were in the treatment group and 65 students were in the control group.

Measured covariates for model 8 include: grade level, attendance rate, FRM status, MAP math percentile scores from September of both the current and prior year, MAP reading percentile scores from September of both the current and prior year and math ability level. In the analysis of model 8, results of tests of covariate balance indicated good balance between groups in both samples, as shown in Table XVI. All variance ratios, except for attendance rate, for matched covariates fall between .5 and 2.0. In addition, the absolute value of standardized differences for all matched variables is less than .2. For example, the standardized match difference for current year September MAP math percentile scores is -.03 and the standardized match difference for prior year September MAP reading percentile scores is -.04. The matched standardized difference for all covariates was also substantially lower than the raw standardized difference, which indicates that the covariates were more comparable following the test application than prior to the test application.

Table XVI

Covariate balance summary for model 8

Covariate	Std. Differences		Variance ratio	
	Raw	Match	Raw	Match
Grade level	-.21	-.14	.97	.97
Att. Rate	.18	-.01	1.90	2.97
FRM	-.13	.03	.60	1.11
CY MAP math percentile score Sept.	1.02	-.03	.71	1.57
CY MAP read percentile score Sept.	.13	-.09	.74	.54
PY MAP math percentile score Sept.	.83	.03	.31	.54
PY MAP read percentile score Sept.	.25	-.04	1.02	1.19

The treatment group for model 9 includes all high-ability white students who were enrolled in an advanced-level math class during the current academic year. The control group for model 9 includes all high-ability white students who were enrolled in grade-level classes. Initial PS matching procedures for model 9 resulted in a total of 307 students matched for the test of average treatment effects for May MAP math percentile scores. 299 students were in the treatment group and only 8 students were in the control group. The small sample size of the control group is due to the small number of white students who are identified as high ability in mathematics and not enrolled in an advanced-level math course. Measured covariates for model 9 include: grade level, attendance rate, FRM status, MAP math percentile scores from September of both the

current and prior year, MAP reading percentile scores from September of both the current and prior year and math ability level.

The PS match for model 9 generated statistically insignificant results. In addition, the results of tests of covariate balance showed weak balance between matched covariates. As a result, the covariate matched statistics are not reported in these findings.

Caliper adjustments and matching with replacement. For models 7 and 8, treatment effects were estimated using matching with narrow calipers, calculated at .25 propensity-score standard deviation units. For model 9, treatment effects were estimated using matching with no caliper. Caliper adjustments for models 7 and 8 ensured a stronger post-match covariate balance. For model 8, matching was conducted with replacement, allowing an untreated student to be matched up to three times. For models 7 and 9, matching was conducted with no replacement. Matching with replacement was used in order to ensure a stronger post-match covariate balance (Caliendo & Kopeinig, 2008).

Logistic regression results. Results for the logistic regression analyses conducted for models 7 through 9 are listed below, in Table XVII. Current year September MAP math percentile scores have a statistically significant relationship with the likelihood of being enrolled in the treatment group across all three models. For models 7 and 8, grade level and prior year September MAP math percentile scores have a statistically significant relationship with the likelihood of being enrolled in the treatment group. Attendance rate also has a statistically significant relationship with the likelihood of being in the treatment group, across models 7 and 8.

For model 7, the *R* squared value is .49, which indicates that the combined predictor variables explain approximately 49 percent of the variance in the likelihood of white students enrolling in an advanced math course. For model 8, the *R* squared value is .28, which indicates that the combined predictor variables explain approximately 28 percent of the variance in the likelihood of enrolling in an advanced math course for white students of average ability. For model 9, the *R* squared value is .31, which indicates that the combined predictor variables explain approximately 311 percent of the variance in the likelihood of enrolling in an advanced math course for high-ability white students.

Table XVII

Logistic Regression Results for Treatment Variables 7-9

Predictor	Treatment variable 7		Treatment variable 8		Treatment variable 9	
	Coef. (β)	Std. Error	Coef. (β)	Std. Error	Coef. (β)	Std. Error
CY grade level	-1.16**	.38	-.95*	.42		
CY att. rate	11.39**	4.12	8.05	5.05	27.10*	12.03
FRM	.80	1.00	.70	.93		
CY MAP math percentile score Sept.	.07***	.02	.06***	.02	.07*	.03
CY MAP read percentile score Sept.	-.02	.01	-.02	.02	.01	.06
PY MAP math percentile score Sept.	.06***	.02	.04**	.02	.03	.07
PY MAP read percentile score Sept.	-.00	.02	-.01	.02	.02	.06
ELA course level	1.40**	.50	1.69**	.58	.74	1.26
R squared	.49		.28		.31	

$***p < .001, **p < .01, * p < .05$

Average treatment effects. Data from the PS analysis show that the average treatment effect was statistically significant for model 8. As mentioned previously, however, there was serious imbalance in pretest math test covariates for models 7 and 9, and this renders the estimation of treatment effects from these models to be problematic when considering the validity of these findings.

As shown in Table XVIII, data indicate that participation in an advanced-level math course had a statistically significant impact on the student achievement outcome variable for white students of average ability. For average-ability white students, participating in an advanced-level math course is associated with an increase of 7.85 percentile points ($p < .01$). As previously mentioned, the test results for model 9 were not statistically significant.

Table XVIII

Estimated effects of treatment group participation on student achievement outcomes for treatment variables 8 and 9

Model	n	Coef.	*SE*	Effect Size *(d)*
Model 8	154	7.85*	3.33	.50
Model 9	307	15.82	11.13	1.75

***$p < .001$, *$p < .05$

CHAPTER V

SUMMARY, DISCUSSION AND RECOMMENDATIONS

This chapter begins with a summary of the study results related to each research question. Following this summary, the chapter provides a more in-depth discussion of these findings, with a focus on how this study fits into the context of prior literature. In addition, this chapter discusses the implications of this research on practitioners and educational leaders. The chapter also describes limitations of this study and recommendations for further research. The chapter ends with a brief conclusion of the study.

Summary of the Findings

The purpose of this study was to explore whether and how tracking structures in mathematics courses at the middle school level relate to differences in achievement between white and black students. This study accomplished this goal by comparing the achievement outcomes of students enrolled in advanced mathematics classes, with students of comparable ability enrolled in grade-level math classes. Specifically, this study used propensity score matching to match students across nine treatment variables to determine whether advanced math course enrollment is statistically significantly associated with increases in student achievement outcomes. The spring MAP assessment in mathematics served as the outcome variable for this study.

Results from the propensity score matching process show that, for average-ability students, enrollment in an advanced-math course is associated with statistically significant improvement in math achievement, as compared to non-treated students. These findings are best supported by data from treatment variables 2, 5 and 8. For each of these variables, the average treatment effect coefficient was found to be statistically significant. Post-match measured covariate tests also indicate a strong balance for pre-test MAP math scores, across all three of these treatment variables. Thus, this finding indicates that average-ability students in advanced math classes do better over time in terms of math achievement then comparable-ability peers who do not participate in the treatment group. The effect size for treatment variables 2, 5 and 8, was moderate which indicates that the propensity score matching results are both practically and statistically significant.

Results from the study also show that increases in student achievement associated with average-ability black student enrollment in advanced-level math courses surpass the increases in math achievement outcomes associated with average-ability white student enrollment in advanced-level math courses. This conclusion is best supported when comparing data from models 5 and 8. For each of these variables, the average treatment effect coefficient was found to be statistically significant. For average-ability black students, participation in an advanced-level course is associated with a percentile score increase of 9.89, whereas average-ability white student participation in the treatment group is only associated with an increase of 7.85 percentile points.

It is these average-ability students who often face the choice of whether to enroll in grade-level versus advanced courses, and these findings suggest that they would be

better served by the latter. These findings have important equity implications because, in the study district, there are more average-ability black than white students, and they are placed or opt to enroll in grade-level math at a higher rate. Because average-ability black student enrollment in advanced-level math courses is associated with a higher percentile score increase than average-ability white student enrollment, the findings suggest that if more average-ability black *and* white students enrolled in advanced-level math, not only would their math achievement improve, but it would reduce the racial math achievement gap in the district.

Results from models 1 and 4 show that, overall and for all black students, enrollment in advanced course enrollment is associated with improvement in student achievement outcomes. Across both of these treatment variables, participation in the treatment group is associated with a statistically significant improvement in achievement on the spring MAP math assessment. Post-match measured covariate tests for these two models also indicate a strong balance for pre-test MAP math scores. This finding shows that the propensity score matching methodology was able to closely match students from the treatment and control groups based on their prior year MAP math scores for these two models. Thus, participation in an advanced math course had a significant impact on achievement outcomes for all students of comparable math ability and for all black students of comparable math ability. In addition, the effect size for models 1 and 4 was moderate which indicates that, the propensity score matching results are both practically and statistically significant.

This study proved inconclusive in determining the impact of advanced-level course enrollment for high-ability math students. According to model 3, the average

treatment effect for all high-ability students enrolled in an advanced-level math course was statistically significant. Nevertheless, the covariate balance between matched pairs for treatment variable 3 was weak. In addition, propensity score matching yielded statistically insignificant results for models 6 and 9, which includes high-ability black and white students, respectively. These findings are not particularly surprising given that most high-ability students opt for advanced-level courses. Thus, it may be difficult to find adequate matches for these students among the enrollees in grade-level courses.

Several factors contributed to these inconclusive results. First, in this study, depending on the treatment variable, high-ability students presented a small sample size in either the treatment or control group. In addition, there was a large discrepancy between high-ability black and white student participation in the treatment group. Black students of high ability were disproportionately enrolled in grade-level math classes, while nearly all white students of high ability were enrolled in advanced-level math classes. Consequently, it is difficult to find adequate matches for these students among the enrollees in grade-level courses.

Though socioeconomic status was not the focus of this study, it was included as a covariate measure in the propensity score matching model. According to the logistic regression results, socioeconomic status, as measured by FRM eligibility, had a statistically significant relationship with the likelihood of treatment course participation for models 1 and 2. For the remaining treatment variables, however, FRM eligibility was not shown to have a statistically significant relationship with likelihood of treatment course participation. Covariate balance tests for FRM across matched pairs were moderate to strong, for students of all ability levels and for average-ability students. This

finding shows that advanced course enrollment has an impact on student math achievement for students of similar socioeconomic status.

Discussion

This study was undertaken with the goal of exploring factors that contribute to inequalities between black and white student achievement outcomes. In addition, this study aimed to further explore the impact of tracking on academic outcomes in mathematics. Findings from this study provide evidence to the scholarly literature in both of these areas and shed additional light on the enduring disparities that exist in the American educational system.

This study reinforces the findings of previous literature related to the impact of tracking on student achievement outcomes in mathematics. Specifically, this study corroborates research that shows that advanced-level course enrollment is associated with a positive impact on a student's academic achievement outcomes. This study also supports earlier findings that determined that average-ability students perform better than peers of similar ability when placed in advanced-level math classes (Mason et al., 1992; Veldman & Sanford, 1984).

Prior research shows that black students are less likely to have access to high quality educational experiences and instructors than their white counterparts, a phenomenon that some scholars call the opportunity gap (Flores, 2007; Harris & Herrington, 2006). This study supports these assertions. For example, in this study, black students of average ability outnumbered white students of average ability by nearly 50 percent. Yet only 73 out of 305 average-ability black students, or 23.93 percent, were enrolled in advanced-level math courses. By contrast, 90 out of 156 average-ability

white students were enrolled in the treatment course, for a total of 57.69 percent. As this study's findings demonstrate, this discrepancy in enrollment has significant effects on student achievement outcomes, since enrollment in advanced math courses is associated with higher student achievement.

Evidence from this study aligns with prior research conducted in the Shaker Heights City School District. Galletta (2013) and Ferguson et al., (2001) emphasized that disparities in advanced-level and honors-course enrollment are partially, if not largely, responsible for gaps in student achievement. Yet this study's results goes beyond the findings of these previous studies in Shaker Heights, by providing rigorous quantitative evidence to show that average-ability black students benefit from advanced-level course enrollment. This study, therefore, suggests how disparities in student achievement outcomes would narrow if disproportionalities in advanced-level math course enrollment were reduced.

Previous research on the achievement gap has produced mixed findings in terms of whether and to what extent discrepancies in academic performance between black and whites narrow, persist or expand as children progress through school (Clotfelter et al., 2009; Fryer & Levitt, 2006; Jencks & Phillips, 1998). Findings from this analysis suggest that tracking structures contribute to growing gaps in academic achievement between black and white students over time at this study site. These gaps expand at least partly because white students of comparable ability to black students are more likely to be enrolled in advanced-level math courses than their black counterparts.

Previous research has extensively documented the effect of a long history of oppression on black-student academic achievement outcomes (Barton & Coley, 2010;

Jencks & Phillips, 1998; Ladson-Billings, 2006). This study reinforces the uphill battle that school districts face to attain equality and equity between racial subgroups. According to the demographic data collected in this study, only 51 black students across the entire study sample of over 1500 students met the target to be considered high-ability math students. This number compares with a total of 307 white students who were identified as having high-mathematics ability across the entire study sample. In stark contrast, only five white students out of the entire study sample were deemed to have low ability in mathematics, compared to 71 black students. These substantial discrepancies serve as a realization of the many historical challenges blacks have faced in America. Ultimately, therefore, this study's data show that black middle-school students in Shaker Heights have two distinct disadvantages compared to their white peers: black students start at a substantially lower-ability level in mathematics than their white peers and are also less likely to have opportunities to improve their academic achievement outcomes as white students of comparable ability.

The Shaker Heights City School District offered a unique and conducive setting to study school tracking structures and the achievement gap. Furthermore, the propensity score model used to analyze this study's data provided quantitative evidence of how tracking contributes to differences in academic outcomes for students of comparable ability. Consequently, this study contributes additional understanding on mathematics education and the achievement gap, and provides additional evidence of the pitfalls of ability grouping in contemporary educational settings.

Implications and Recommendations for Practice

Findings from this study contain implications for practitioners in the field as they strive for equity and rigorous learning experiences in the field of math instruction and in education overall. This study focused on specific grade levels in a specific school district, with characteristics that may distinguish it from other school settings in this country and abroad. In addition, this study used propensity score analysis, which demonstrated correlation between the measured covariates of this study and the achievement outcome. When considering the implementation of recommendations from this study, therefore, practitioners should bear in mind the confines of this study and be cognizant that different contextual settings may yield different results. Nevertheless, reasonable implications and recommendations for practitioners do emerge from the results of this study, and the data on which they are based.

On a broad level, this study reinforces the notion that student achievement outcomes improve when all students are challenged with rigorous and engaging curriculum and courses. More specifically, findings from this study underscore that additional efforts should be made to expand the inclusion of black students in advanced-level courses. In particular, school district leaders and teachers should focus on expanding the participation of average-ability black students in advanced-level math classes. Two findings from this study support this approach. First, this study confirms the hypothesis that, in general, enrollment in advanced-level courses is associated with improvements in math achievement. Secondly, this study demonstrates that both white and black students of average ability benefit from being enrolled in higher-level math courses.

The scope of this study falls short of providing a conclusive recommendation regarding the benefits or drawbacks of broader structural reforms, such as the elimination of middle-school level math tracking in its entirety or the universal acceleration of students into advanced-level courses (Allensworth et al., 2009; Domina et al., 2014; Domina & Saldana, 2012). Nevertheless, this study corroborates previous research that shows that tracking is connected to disparities in achievement (Oakes, 2005). Practitioners should take these findings into consideration when making decisions about math course leveling.

As previous literature has shown, changing tracking structures alone, without simultaneously strengthening classroom instruction and building teacher efficacy will prove insufficient in altering the trajectory of student achievement outcomes (Bernhardt, 2014; Slavin, 1990; Watanabe, 2007). Findings from this study should therefore be shared with teachers and students, so that they can help take the lead in considering alternatives to ability grouping and in continuing to research the effects of tracking on achievement outcomes (Bernhardt, 2014; Watanabe, 2007). In this way, data from this study can ultimately boost efforts to further explore and address underlying inequities in the American educational system.

Finally, this study highlights the disproportionately small number of middle-school black students who have demonstrated a high ability in mathematics. While not the primary focus of this study, these findings underscore the need to provide additional supports and instructional interventions for black students in mathematics from an earlier age and on a broader scale, in order to ensure a more proportional representation of black students at higher-ability levels. For example, policy leaders must consider the

implementation of larger-scale societal reforms to address inequities in housing policies, social welfare supports and access to a high standard of living that contribute to the substantial gap in academic achievement.

Limitations of the Study

This study provided robust evidence regarding the impact of tracking on middle-school student achievement in mathematics. Nevertheless, there are limitations of this analysis that are inherent to the study design and methodology. Consequently, some caution should be applied when making causal interpretations based on this study.

This study used a quasi-experimental design to ensure that treatment and control groups were comparable based on observed measured covariates. These measured covariates were selected based on prior research and to encompass a broad range of potential contributing factors. As noted earlier in the study, however, additional unobserved characteristics may exist that influence placement in treatment groups or that impact academic achievement outcomes. For example, factors such as internal student motivation and student home backgrounds were beyond the scope of this study. In addition, selection bias can contribute to differences in control and treatment group populations in studies such as this one (Coca et al., 2012).

The propensity score matching methodology of this study partially addresses these concerns by including both baseline data and demographic data and by including post-match balance tests to help assure robustness of the study findings (Steiner et al., 2010). Nevertheless, suboptimal covariate balance serves as a limitation of this study across some models or covariate matches. For example, post-match balance tests in some models attained weaker balance for pretest reading percentile scores than with pretest

math percentile scores. More importantly, this study's models failed to adequately match covariates for models comprised of high-ability students. In these high-ability models, post-match balance tests showed that the PSM model achieved weak and inadequate balance across most, if not all, covariates and particularly in math pretest data. These covariate balance limitations must be taken into consideration when attempting to generalize the findings of this study and thus, scholars should be hesitant to apply the findings of this study to these specific subgroups of students.

Nonetheless, when reviewing this study, educational leaders and practitioners should be concerned with the underlying causes of this limitation. In this study, out of 87 high-ability black students, only 56.32 percent, or 49, were enrolled in an advanced-level math class. By contrast, 96.17 percent of high-ability white students, or 326 out of 339 total students, were enrolled in an advanced-level math class. These data demonstrate that in order to fully understand the effects of tracking on high-ability math students, high-ability black students must be enrolled in advanced-level courses at a comparable rate – or at least at a minimum enough proportion in order to be statistically significant – as their white peers. Ultimately, therefore, this study reinforces the concept of an opportunity gap highlighted in earlier research (Flores, 2007).

Cases were also discarded in this study due to failure to match within a caliper. Excluded students were shown to have significantly higher rates of FRM eligibility, or significantly lower attendance rates, MAP math percentile scores, MAP reading percentile scores, or lower English course levels. Some students were also not included in the tests of average treatment effects because of the small sample size of a particular subgroup within the overall study population. These subgroups included students with

161

Individualized Education Plans (IEPs), students who were enrolled in accelerated math courses two grade levels above their current grade level and English Language Learners. These exclusions should be taken into account when considering the generalizability of this study's results. Specifically, this study's findings may not be applicable to students who present similar characteristics as the excluded subgroups.

Further Study

Results from this study point the way toward several areas for future research. For example, additional research continues to be warranted in regard to students of high math ability. Prior research has equivocated in terms of the impact of tracking on achievement outcomes for students of high ability, and this study's findings were inconclusive in contributing to this discussion. In particular, there are gaps in understanding for how tracking affects the academic outcomes of high-ability black students. These gaps in scholarly research can be partly attributable to smaller sample sizes of high-ability black math students. Thus, future studies must aim to incorporate larger study populations with a specific focus on this subgroup.

Data from this study also showed that, in certain cases, prior achievement in reading statistically significantly correlates with academic achievement outcomes in tracked mathematics settings. Further studies should explore this interaction in more depth. Future research should also continue to attempt to identify some of the other observable and unobservable covariates that lead to disparities in academic achievement outcomes in tracked settings.

Conclusion

This study was designed to investigate the relationship between tracking structures in mathematics and student achievement outcomes at the middle school level. This relationship was explored through a series of propensity score matching tests and using a quasi-experimental design. Results from this study indicate that participation in an advanced-level math course is associated with improvements in math achievement outcomes. Furthermore, this study showed that average-ability black and white students benefit from enrollment in higher-level math classes.

On a broader scale, this study also aimed to further probe the fault lines that are engrained in the American educational system. This study further substantiates the assertions of previous scholars, who have determined that the practice of ability grouping undermines the foundation of the American democratic framework and, "in direct violation of the tenets of the American dream, keep[s] schools from helping all students to pursue individual success" (Hochschild & Scovronick, 2003, p. 162). School districts and educational leaders are encouraged to look at this study for recommendations to make progress in narrowing the achievement gap and in reducing societal disparities in academic and lifelong outcomes. As scholars continue to explore the challenges and dilemmas associated with tracking and the black-white achievement gap, this study's findings should continue to guide the work so that ultimately, all students, and the American society at large, can benefit from an equitable and high quality educational experience.

REFERENCES

Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica, 74*(1), 235-267.

Abadie, A., Dukker, D. M., Herr, J. L., & Imbens, G. W. (2004). Implementing matching estimators for average treatment effects in Stata. *Stata Journal, 4,* 290-311.

Adams, D. W. (1995). *Education for extinction: American Indians and the boarding school experience, 1875-1928.* Lawrence, KS: University Press of Kansas.

Akos, P., Shoffner, D. M., & Ellis, D. M. (2007). Mathematics placement and the transition to middle school. *Professional School Counseling, 10*(3), 238-244.

Alexander, K. L., & Cook, M. A. (1982). Curricula and coursework: A surprise ending to a familiar story. *American Sociological Review*, 626-640.

Allensworth, E., Nomi, T., Montgomery, N., & Lee, V. E. (2009). College preparatory curriculum for all: Academic consequences of requiring algebra and English I for ninth graders in Chicago. *Educational Evaluation and Policy Analysis, 31*(4), 367-391.

Alvarez, D., & Mehan, H. (2006). Whole-school detracking: A strategy for equity and excellence. *Theory into practice, 45*(1), 82-89.

Anderson, S., Medrich, E., & Fowler, D. (2007). Which achievement gap? *Phi Delta Kappan, 88*(7), 547-550.

Argys, L. M., Rees, D. I., & Brewer, D. J. (1996). Detracking America's schools: Equity at zero cost? *Journal of Policy analysis and Management*, 623-645.

Ayres, L. P. (1913). *Laggards in our schools: A study of retardation and elimination in city school systems.* New York, NY: Survey Associates.

Balfanz, R., & Byrnes, V. (2006). Closing the mathematics achievement gap in high-poverty middle schools: Enablers and constraints. *Journal of Education for Students Placed at Risk, 11*(2), 143-159.

Barton, P. E., & Coley, R. J. (2010). The black-white achievement gap: When progress stopped. Policy Information Report. *Educational Testing Service.*

Bernhardt, P. E. (2014). How do I get in? Criteria shaping the high school course recommendation process. *Current Issues in Education, 17*(1).

Boaler, J., & Sengupta-Irving, T. (2016). The many colors of algebra: The impact of equity focused teaching upon student learning and engagement. *The Journal of Mathematical Behavior.*

Boaler, J., & Staples, M. (2008). Creating mathematical futures through an equitable teaching approach: The case of Railside School. *Teachers College Record, 110*(3), 608-645.

Bohan, C. H. (2003). Early vanguards of progressive education: The Committee of Ten, the Committee of Seven, and social education. *Journal of Curriculum & Supervision, 19*(1), 73.

Bohrnstedt, G., Kitmitto, S., Ogut, B., Sherman, D., & Chan, D. (2015). School composition and the black-white achievement gap. NCES 2015-018. *National Center for Education Statistics.*

Brazer, S.D. (2018). Shaker Heights Gifted Education Workgroup data slides (Unpublished). Shaker Heights City School District.

Burris, C. C. (2014). *On the same track: How schools can join the twenty-first-century struggle against resegregation.* Boston, MA: Beacon Press.

Burris, C. C., & Welner, K. G. (2005). Closing the achievement gap by detracking. *The Phi Delta Kappan*, *86*(8), 594-598.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, *22*(1), 31-72.

Carpenter, D. M., Ramirez, A., & Severn, L. (2006). Gap or gaps: Challenging the singular definition of the achievement gap. *Education and Urban Society*, *39*(1), 113-127.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2009). The academic achievement gap in grades 3 to 8. *The Review of Economics and Statistics*, *91*(2), 398-419.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2015). The aftermath of accelerating algebra evidence from district policy initiatives. *Journal of Human Resources*, *50*(1), 159-188.

Coates, T. N. (2015). The black family in the age of mass incarceration. *The Atlantic*, *316*(3), 82.

Coca, V., Johnson, D., Kelley-Kemple, T., Roderick, M., Moeller, E., Williams, N., & Moragne, K. (2012). Working to my potential: The postsecondary experiences of CPS students in the International Baccalaureate Diploma Programme. *Consortium on Chicago School Research*.

Cohen, E. G., & Lotan, R. A. (1997). *Working for equity in heterogeneous classrooms: sociological theory in practice. Sociology of Education Series*. New York, NY: Teachers College Press.

Davis, T. M. (2014). School choice and segregation "tracking" racial equity in magnet schools. *Education and Urban Society*, *46*(4), 399-433.

Delbanco, A. (2000). *The real American dream*. Harvard University Press.

Domina, T., & Saldana, J. (2012). Does raising the bar level the playing field? Mathematics curricular intensification and inequality in American high schools, 1982–2004. *American Educational Research Journal*, *49*(4), 685-708.

Domina, T., McEachin, A., Penner, A., & Penner, E. (2015). Aiming high and falling short: California's eighth-grade algebra-for-all effort. *Educational Evaluation and Policy Analysis*, *37*(3), 275-295.

Dougherty, S. M., Goodman, J. S., Hill, D. V., Litke, E. G., & Page, L. C. (2015). Middle school math acceleration and equitable access to eighth-grade algebra: Evidence from the Wake County Public School System. *Educational Evaluation and Policy Analysis*, *37*(1_suppl), 80S-101S.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY: Random House Incorporated.

Eyler, J. et al. (1982). Resegregation: Segregation within desegregated schools. *American Educational Research Association*.

Ferguson, R. (1998). Can schools narrow the black-white test score gap? In Jencks, C., & Phillips, M. (Eds.), *The Black-White test score gap* (pp. 318-374). Washington, DC: Brookings Institution Press.

Ferguson, R. F., Ludwig, J., & Rich, W. (2001). A diagnostic analysis of black-white GPA disparities in Shaker Heights, Ohio. *Brookings papers on education policy*, (4), 347-414.

Finley, M. K. (1984). Teachers and tracking in a comprehensive high school. *Sociology of Education*, 233-243.

Flores, A. (2007). Examining disparities in mathematics education: Achievement gap or opportunity gap? *The High School Journal*, *91*(1), 29-42.

Fryer Jr, R. G., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *The Review of Economics and Statistics*, *86*(2), 447-464.

Fryer Jr, R. G., & Levitt, S. D. (2006). The black-white test score gap through third grade. *American Law and Economics Review*, *8*(2), 249-281.

Fuhrman, S., & Lazerson, M. (2005). *The institutions of American democracy: The public schools*. New York, NY: Oxford University Press.

Galletta, A. (2013). *Mastering the semi-structured interview and beyond: From research design to analysis and publication*. New York, NY: NYU Press.

Gamoran, A. (2001). American schooling and educational inequality: A forecast for the 21st century. *Sociology of education*, 135-153.

Gamoran, A., & Hannigan, E. C. (2000). Algebra for everyone? Benefits of college-preparatory mathematics for students with diverse abilities in early secondary school. *Educational Evaluation and Policy Analysis*, *22*, 241–254.

Gamoran, A., & Mare, R. D. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality? *American journal of Sociology*, 1146-1183.

Gilens, M. (2009). Preference gaps and inequality in representation. *PS: Political Science & Politics*, *42*(2), 335-341.

Goldstein, D. (2014). *The teacher wars: A history of America's most embattled profession*. Anchor.

Grant, G. (2009). *Hope and despair in the American city*. Boston, MA: Harvard

   University Press.

Grant, J. (2014). *The boy problem: Educating boys in urban America, 1870–1970*. JHU

   Press.

Guo, S., & Fraser, M. W. (2015). *Propensity score analysis: Statistical methods and*

   applications (2nd ed.). Thousand Oaks, CA: Sage.

Hall, G. E. (1970). On the road to educational failure: A lawyer's guide to tracking.

   Inequality in Education, Harvard Center for Law and Education, 5, 1-27.

Hallinan, M. T. (1994). Tracking: From theory to practice. *Sociology of Education, 67*(2),

   79-84.

Hallinan, M. T. (2004). The detracking movement. *Education Next, 4*(4).

Hanushek, E. A. & Wossman, L. (2006). Does educational tracking affect performance

   and inequality? Differences-in-differences evidence across countries. *The*

   Economic Journal, 116(510).

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the

   assessment of measured covariate balance to test causal associations in

   psychological research. *Psychological methods, 15*(3), 234.

Harris, D. N., & Herrington, C. D. (2006). Accountability, standards, and the growing

   achievement gap: Lessons from the past half-century. *American Journal of*

   Education, 112(2), 209-238.

Hartman, A. (2015). *A war for the soul of America: A history of the culture wars*.

   Chicago, IL: University of Chicago Press.

Heyns, B. (1974). Social selection and stratification within schools. *American Journal of Sociology, 79*(6), 1434-1451.

Hochschild, J. L., & Scovronick, N. (2003). *The American dream and the public schools*. New York, NY: Oxford University Press.

Hoffer, T. B. (1992). Middle school ability grouping and student achievement in science and mathematics. *Educational evaluation and policy analysis, 14*(3), 205-227.

James Coleman et al., (1966). *Equality of educational opportunity*. Washington, D.C.: U.S. Government Printing Office.

Jencks, C., & Phillips, M. (Eds.). (1998). *The black-white test score gap*. Washington, DC: Brookings Institution Press.

Kaestle, C. (1983). *Pillars of the republic: Common schools and American society, 1780-1860*. New York, NY: Hill and Wang.

Kane, T. J., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic perspectives, 16*(4), 91-114.

Kao, G., & Thompson, J. S. (2003). Racial and ethnic stratification in educational achievement and attainment. *Annual review of sociology*, 417-442.

Kulik, C. L. C., & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American educational research journal, 19*(3), 415-428.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in US schools. *Educational researcher, 35*(7), 3-12.

Lareau, A. (2000). *Home advantage: Social class and parental intervention in elementary education*. Lanham, MD: Rowman & Littlefield Publishers.

Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity? *Educational researcher*, *31*(1), 3-12.

Leow, C., Marcus, S., Zanutto, E., & Boruch, R. (2004). Effects of advanced course-taking on math and science achievement: Addressing selection bias using propensity scores. *American Journal of Evaluation*, *25*(4), 461-478.

Linchevski, L., & Kutscher, B. (1998). Tell me with whom you're learning, and I'll tell you how much you've learned: Mixed-ability versus same-ability grouping in mathematics. *Journal for Research in Mathematics Education*, 533-554.

Loveless, T. (2009). Tracking and detracking: High achievers in Massachusetts middle schools. *Thomas B. Fordham Institute*.

Loveless, T. (2011). *The tracking wars: State reform meets school policy*. Washington, DC: Brookings Institution Press.

Mason, D. A., Schroeter, D. D., Combs, R. K., & Washington, K. (1992). Assigning average-achieving eighth graders to advanced mathematics classes in an urban junior high. *The Elementary School Journal*, *92*(5), 587-599.

McGuinn, P. (2016). From no child left behind to the every student succeeds act: Federalism and the education legacy of the Obama administration. *Publius: The Journal of Federalism*, *46*(3), 392-415.

Meehan, Thomas (1963, March). The good life in Shaker Heights. *Cosmopolitan Magazine*, 46-51.

Molyneaux, D., & Sackman, S. (1987). '75 years: An informal history of Shaker Heights. *Shaker Heights Public Library, Shaker Heights, OH*.

Moore, J. R. (2004). 50 years after Brown: Segregation in the Miami-Dade County public

    schools. *Equity & Excellence in Education*, *37*(3), 289-301.

Moses, R., Kamii, M., Swap, S. M., & Howard, J. (1989). The algebra project:

    Organizing in the spirit of Ella. *Harvard Educational Review*, *59*(4), 423-444.

Moynihan, D. P. (1965), *The negro family: The case for national action*, Office of Policy

    Planning and Research, United States Department of Labor, Washington, DC.

NAEP Assessment data. Retrieved from https://www.nationsreportcard.gov/

National Assessment of Educational Progress website

    (https://www.nationsreportcard.gov)

National Council of Teachers of Mathematics. Commission on Standards for School

    Mathematics. (1989). *Curriculum and evaluation standards for school*

    mathematics. National Council of Teachers of Mathematics.

National Education Association of the United States. Commission on the Reorganization

    of Secondary Education. (1918). *Cardinal principles of secondary education: A*

    report of the commission on the reorganization of secondary education (No. 35).

    Government printing office.

National Education Association of the United States. Committee of Ten on Secondary

    School Studies. (1894). *Report of the committee of ten on secondary school*

    studies: With the reports of the conferences arranged by the committee. National

    Education Association.

National Middle School Association. (2003). *This we believe: Successful schools for*

    young adolescents: A position paper of the National Middle School Association.

    National Middle School Association.

Northwest Evaluation Association (2011). *Technical manual for Measures of Academic Progress (MAP) and Measures of Academic Progress for Primary Grades (MPG)*. Retrieved from https://community.nwea.org.

Northwest Evaluation Association (2016). *Linking the Ohio state assessments to NWEA MAP Growth tests*. Portland, OR: NWEA.

Oakes, J. (1992). Can tracking research inform practice? Technical, normative, and political considerations. *Educational Researcher*, *21*(4), 12-21.

Oakes, J. (1997). Detracking: The social construction of ability, cultural politics, and resistance to reform. *Teachers College Record*, *98*(3), 482-510.

Oakes, J. (2005). *Keeping track*. New Haven, CT: Yale University Press.

Oakes, J., Ormseth, T., Bell, R., & Camp, P. (1990). Unequal opportunities: The effects of race, social class, and ability grouping on access to science and mathematics education. *Santa Monica, CA: The RAND Corporation*.

Ogbu, J. U. (2003). *Black American students in an affluent suburb: A study of academic disengagement*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Parsons, T., & Halsey, A. H. (1959). The school class as a social system. *Schools and society: A sociological approach to education*, 32-40.

Pourdavood, R. G., Cowen, L. M., Svec, L. V., Skitzki, R., & Grob, S. (1999). A paradoxical path to reform: The case study of Lomond Elementary School. Transforming Learning Communities. *Ohio Department of Education*.

Putnam, R. D. (2015). *Our kids: The American dream in crisis*. New York, NY: Simon and Schuster.

Rampey, B.D., G.S. Dion, and P.L. Donahue (2009), *NAEP 2008 trends in academic progress*, National Center for Education Statistics, Institute of Educational Sciences, U.S. Department of Education, Washington, D.C.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither opportunity*, 91-116.

Reardon, S. F., & Galindo, C. (2009). The Hispanic-white achievement gap in math and reading in the elementary grades. *American Educational Research Journal*, *46*(3), 853-891.

Reese, W. J. (2005). *America's public schools: from the common school to" No Child Left Behind"*. Baltimore, MD: JHU Press.

Rice, L., Barth, J. M., Guadagno, R. E., Smith, G. P., & McCallum, D. M. (2013). The role of social support in students' perceived abilities and attitudes toward math and science. *Journal of youth and adolescence*, *42*(7), 1028-1040.

Roscigno, V. J. (1998). Race and the reproduction of educational disadvantage. *Social Forces*, *76*(3), 1033-1061.

Roscigno, V. J., & Ainsworth-Darnell, J. W. (1999). Race, cultural capital, and educational resources: Persistent inequalities and achievement returns. *Sociology of education*, 158-178.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33-38.

Rosenbaum, P.R. (2002). *Observational studies* (2nd ed.). New York, NY: Springer-Verlag.

Rothstein, R. (2017). *The color of law: A forgotten history of how our government segregated America.* New York, NY: Liveright Publishing.

Rubin D.B. & Thomas N. (1996). Matching using estimated propensity scores: relating theory to practice. *Biometrics, 52,* 249–264.

Rui, N. (2009). Four decades of research on the effects of detracking reform: Where do we stand?—A systematic review of the evidence. *Journal of evidence-based medicine, 2*(3), 164-183.

Ryan, J., Barnes, L.H., & Torres, A. (2018). *Four years strong in San Francisco: Holding an equity-based detracking policy over time.* San Francisco Unified School District.

Sahm, C. (2015). What explains success at Success Academy? *Education Next, 15*(3).

Saporito, S., & Sohoni, D. (2007). Mapping educational inequality: Concentrations of poverty among poor and minority students in public schools. *Social Forces, 85*(3), 1227-1253.

Sawchuck, Stephen (2018, June, 12). A bold effort to end algebra tracking shows promise. *Education Week, 37*(35), 1, 11.

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational design.* American Educational & Research Association.

Senechal, J. (2014). Middle grades math: Assessing the debate over the when and how of algebra. Retrieved from http://scholarscompass.vcu.edu/merc_pubs/9

Shaker Heights City School District (2018). *2017-2019 Academic Planning Guide*.

    Retrieved from http://www.shaker.org/academicplanningguide.aspx.

Shaker Heights City School District enrollment summary report, Retrieved November,

    2017.

Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A

    best-evidence synthesis. *Review of educational research*, *60*(3), 471-499.

Slavin, R. E., & Madden, N. A. (2006). Reducing the gap: Success for all and the

    achievement of black students. *The Journal of Negro Education*, 389-400.

Smeeding, T. M. (2005). Public policy, economic inequality, and poverty: The United

    States in comparative perspective. *Social Science Quarterly*, *86*(s1), 955-983.

Smith, D., Frey, N., Pumpian, I., & Fisher, D. (2017). *Building equity: Policies and*

    practices to empower all learners. Alexandria, VA: ASCD.

Southworth, S., & Mickelson, R. A. (2007). The interactive effects of race, gender and

    school composition on college track placement. *Social Forces*, *86*(2), 497-523.

Steele, C. M. (2011). *Whistling Vivaldi: How stereotypes affect us and what we can do*.

    New York, NY: WW Norton & Company.

Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of

    covariate selection in controlling for selection bias in observational studies.

    Psychological methods, *15*(3), 250.

Sterling, R. W., & Scott, W. C. (1996). *Plato: The Republic*. New York, NY: WW

    Norton & Co.

Stinson, D. W. (2004). Mathematics as "gate-keeper" (?): Three theoretical perspectives

    that aim toward empowering all children with a key to the gate.

Stone, C. A., & Tang, Y. (2013). Comparing propensity score methods in balancing covariates and recovering impact in small sample educational program evaluations. *Practical Assessment, Research & Evaluation*, *18*(13).

Thernstrom, A., & Thernstrom, S. (2003). *No excuses: Closing the racial gap in learning*. New York, NY: Simon and Schuster.

Thomas, J. Y., & Brady, K. P. (2005). Chapter 3: The Elementary and Secondary Education Act at 40: Equity, accountability, and the evolving federal role in public education. *Review of Research in Education*, *29*(1), 51-67.

Thum Y. M., & Hauser, C. H. (2015). *NWEA 2015 MAP norms for student and school achievement status and growth*. NWEA Research Report. Portland, OR: NWEA.

Thum, Y. M., & Matta, T. (2015). *MAP college readiness benchmarks*. NWEA Research Report. Portland, OR: NWEA.

Trujillo, T. M. (2012). The politics of district instructional policy formation: Compromising equity and rigor. *Educational Policy*, 0895904812454000.

US Census Bureau (2016). Quick facts: Shaker Heights City, Ohio. https://www.census.gov/.

Useem, E. L. (1992). Middle schools and math groups: Parents' involvement in children's placement. *Sociology of education*, 263-279.

Veldman, D. J., & Sanford, J. P. (1984). The influence of class ability level on student achievement and classroom behavior. *American Educational Research Journal*, *21*(3), 629-644.

Vinovskis, M. A. (2001). Overseeing the nation's report card: The creation and evolution of the National Assessment Governing Board (NAGB).

Voight, A., & Velez, V. (2018). Youth participatory action research in the high school

    curriculum: Education outcomes for student participants in a district-wide

    initiative. *Journal of Research on Educational Effectiveness*, 1-19.

Warner, R. M. (2013). *Applied statistics: from bivariate through multivariate techniques:*

    from bivariate through multivariate techniques (2nd ed.). Thousand Oaks, CA:

    Sage.

Watanabe, M. (2007). Lessons from a teacher inquiry group about tracking: Perceived

    student choice in course-taking and its implications for detracking reform. *The*

    Teachers College Record, *109*(9), 2136-2170.

Waters, M. C., & Eschbach, K. (1995). Immigration and ethnic and racial inequality in

    the United States. *Annual review of sociology*, *21*(1), 419-446.

Welner, K. G. (2001). *Legal rights, local wrongs: When community control collides with*

    educational equity. Albany, NY: SUNY Press.

Yonezawa, S., Wells, A. S., & Serna, I. (2002). Choosing tracks: "Freedom of choice" in

    detracking schools. *American Educational Research Journal*, *39*(1), 37-67.

Young, M. F. (1972). Knowledge and control: New directions for the sociology of

    education. London: Callier, Macmillan.