8-2013

Does Homogenous Ability Grouping for High School Honors English Instruction Benefit the High Achiever?

Douglas Paul Hostetter

Does Homogeneous Ability Grouping for High School Honors English Instruction Benefit the High Achiever?

BY

Douglas Paul Hostetter

Dissertation Committee:

Gerard Babo, Ed.D., Mentor
James Caulfield, Ed.D., Reader
Joseph Potts, Ph.D., Reader

Submitted in Partial Fulfillment
of the Requirements for the Degree
Doctor of Education
Seton Hall University

2013

SETON HALL UNIVERSITY
COLLEGE OF EDUCATION AND HUMAN SERVICES
OFFICE OF GRADUATES STUDIES

APPROVAL FOR SUCCESSFUL DEFENSE

Doctoral Candidate, **Douglas Paul Hostetter,** has successfully defended and made the required modifications to the text of the doctoral dissertation for the **Ed.D.** during the **Summer Term 2013.**

DISSERTATION COMMITTEE
(PLEASE SIGN AND DATE BESIDE YOUR NAME)

Mentor:
Dr. Gerard Babo _____ 6/4/13

Committee Member:
Dr. James Caulfield _____ 6/4/13

Committee Member
Dr. Joseph Potts _____ 06.04.13

The mentor and any other committee members who wish to review revisions will sign and date this document only when revisions have been completed. Please return this form to the Office of Graduate Studies, where it will be placed in the candidate's file and submit a copy with your final dissertation to be bound as page number two.

ABSTRACT

Public schools are examining their policies and instructional practices to address the achievement gap exposed by the reporting requirements of NCLB (Wenglinski, 2004). As accountability measures and stakes rise, there is a call for an improved use of scientific evidence to inform educational policymaking (Wiseman, 2010). In terms of the achievement gap, national studies at the secondary level show when students are grouped according to ability there is a rise in achievement inequality between the groups (Gamoran & Mare, 1989; Hoffer, 1992). The purpose of this study is to determine if there is a tangible, measurable academic benefit to homogeneously grouping high school honors English students in a diverse, suburban school district in Washington State.

The following research question guided this study: To what extent, if any, does ability grouping of high achieving students defined as 9th and 10th grade honors English students in a suburban Washington State school district, affect their performance on state and pre-college assessments of reading and writing achievement when controlling for student mutable variables?

The research design used a non-experimental, explanatory associational design. Student achievement measures were examined before, during, and after high school students were placed in either like-ability (homogeneous) or mixed-ability (heterogeneous) groups when receiving honors English instruction in both their 9th and 10th grade years.

The findings of this study indicate that type of grouping is not always a significant contributor to students' scores on state and pre-college assessments of reading and writing achievement. In models where grouping was a significant contributor, the favored grouping type was not homogeneous but heterogeneous grouping.

Due to the relatively small sample size of this study, it is recommended that additional studies be done to answer the research question. This question is essential in our educational

system's continued pursuit of equality of educational outputs; or in other words, closing and eliminating the achievement gap. It is recommended that a larger-scale experimental, quantitative study be done to determine if the results of this study can be replicated.

ACKNOWLEDGMENTS

My mentor, Dr. Gerard Babo, I want to thank you for being so responsive and accessible during this process. Your guidance and assistance with the writing of this dissertation and the manipulation and analysis of the data via SPSS was invaluable.

I would also like to thank, Dr. James Caulfield for his wisdom, humor and inspiration. You are a model of intellect and grace.

Dr. Joe Potts, thank you for pushing me to "get the thing done." You are a great colleague whose support throughout the process was a great motivator.

DEDICATION

My wife Jayme, thank you for supporting me on this venture, even though it meant you had your entire family in college. I couldn't have done this without you. I love you.

Nathan and Natalie, my wonderful children who I love and am so proud. Continue to challenge your mind and chase your dreams. I hope that Dad has shown you that you are never too old to stop learning in order to achieve your goals.

Mom and Dad, thanks for modeling for me the importance of an education and for all your love and support over these many years.

Jim Oliver, my doctoral companion: thanks for making the process bearable and even enjoyable. I look forward to sharing our many future accomplishments.

D.P.H.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

Introduction

In his Ordinances of 1785 and 1787, Thomas Jefferson proposed a public education system that would level the playing field between those who came from a more privileged background and those who did not have such advantages. He insisted that democracy was the basic theme of education and an educated nation would help society progress (Tanner & Tanner, 2007). In the mid-1800s, Horace Mann extended these Jeffersonian ideals by citing the importance of heterogeneous grouping as a unifying practice in public schools that brought together students of diverse backgrounds. Similarly, John Dewey by the early twentieth century advanced the idea that knowledge should be equally available to all members of society not just the privileged who then make decisions for the rest of society. The ideals of an inclusive or heterogeneous educational system were extended by the *Cardinal Principles* of 1918 and Ralph Tyler's 1930s *Eight Year Study* (Tanner & Tanner, 2007). These two studies set the tone for educational discourse by advocating a flexible curriculum, which considered the needs of the learner. Dewey's progressivist ideology dominated educational discourse until the advent of World War II.

The outbreak of World War II in the 1940s and later the Cold War, stymied the progressivist momentum and that of mixed ability or heterogeneous grouping. These two monumental events launched the American education system into a more discipline-centered focus (Chayte, 2010; Tanner & Tanner, 2007). Indeed, James Conant's influential report The *American High School Today* published in 1959 advocated a return to the practice of grouping students according to ability. Conant (1959) asked the question: can a comprehensive high school educate "those with a talent for handling advanced academic subjects" (p. 15)? Conant

generally found that the academically talented student was not being sufficiently challenged. Conant concluded that students of similar ability should have almost identical programs; "In the required subjects...students should be grouped according to ability, subject by subject" (p. 49).

In the context of the space race and the resultant national pressure to reinvigorate our scientific community, Bruner published *The Process of Education* (1960). Bruner advanced the notion that in the long-range crisis in national security a successful resolution will depend on a well-educated citizenry. According to Bruner, "Excellence must not be limited to the gifted student. But the idea that teaching should be aimed at the average student in order to provide something for everybody is an equally inadequate formula" (p. 70). Although not specifically advocating ability grouping, his work promoted an educational focus on the academically gifted. Together these two reports along with the social and political pressures of the Cold War reestablished the theoretical framework in American education that grouping students according to ability benefits students' learning (Tanner & Tanner, 2007). In the context of the Cold War and beyond, the conceptual framework of grouping for instructional efficiency was reborn (Allan, 1991; Benbow & Stanley, 1996; Gamoran, 2009; Gamoran & Weinstein, 1998; Oakes & Guiton, 1995). Essentialism was rekindled with a focus on advancing the math and science education of our more gifted students.

This essentialist shift in the American educational system occurred in a time of significant social unrest and change. In 1954, the United States Supreme Court in their landmark ruling *Brown v. Topeka Board of Education* held that segregated education was inherently unequal. The Cold War together with the desegregation demands of the 1950s prompted a movement to separate gifted students in order to provide them a special education. Southern states used ability grouping to avoid desegregation orders and the northern cities used ability

grouping in response to the large migration of blacks (Chayte, 2010). In response to *Brown v. Topeka,* the United States Congress enacted the Civil Rights Act of 1964 (CRA, P.L. pp. 88-352). This act provided many protections, among them prohibiting racial segregation in our public schools.

Additionally, in 1965, Congress authorized the Elementary and Secondary Education Act (ESEA) in support of Lyndon Johnson's *war on poverty.* President Johnson emphasized the importance of all children having access to a quality education as a means to leading productive lives. In 1975, Congress passed the Education for All Handicapped Children Act (EHA, P.L. pp. 94-142). This act required federally funded schools to provide a free appropriate public education (FAPE) to students with disabilities. Individually and collectively, these pieces of legislation and the findings of the court had lasting and far-reaching effects on United States policy and practice. In terms of educational effects, public schools were required to desegregate racially and to include all children regardless of the presence of disabilities.

In 1983, the National Commission on Excellence in Education released their report, *A Nation at Risk: The Imperative for Educational Reform.* This report found that many American students were functionally illiterate, SAT scores were trending downward, and students were in need of an increasing number of remedial courses in college. The report concluded that these findings and trends threatened our educational system and our nation's future. In the end, the report served to heighten and solidify federal involvement in public education (Hewitt, 2008). *Goals 2000,* which became law in 1994, furthered the federal government's role by providing funding to states who were implementing school reforms and developing education standards and assessments. Signed into law in 2002, the 2001 No Child Left Behind Act substantially increased the testing requirements and set demanding accountability standards for schools,

districts, and states (Bloomfield, 2003). Concerns gradually arose that state standards varied widely. Comparing student performance on state assessments with their performance on the NAEP clearly demonstrated variability in the rigor of each state's tests (Rothman, 2012).

Due to this variability, content standards are becoming nationalized and are currently referred to as the Common Core State Standards (Rothman, 2012). To assess a school's progress in meeting these common standards, public school districts and their schools were asked to demonstrate that they were making adequate yearly progress (AYP). AYP did not use a broad measure of a school's overall averages but required schools to report on the achievement of a number of student subgroups to determine school effectiveness (Fusarelli, 2004). NCLB has as one of its goals to diminish the achievement gap between minority and nonminority children, especially between low socio-economic students and their more affluent classmates (Bloomfield & Cooper, 2003; Day-Vines & Patton, 2003; Sunderman, 2003).

NCLB's goal of reducing this gap brings concern from some educators, parents, politicians, and business leaders. They believe that if America is to remain internationally competitive the academic potential of our top students must be maximized (Xiang, Dahlin, Cronin, Theaker, & Durant, 2011). In a study of test score gains of students, Kober, McMurrer, and Silva (2011) found that over a seven-year period lower performers increased to a larger degree than did the higher achievers. Researchers and advocates of the high achieving student believe school objectives should ensure that all students maintain an upward trajectory. Therefore, the focus should not disproportionately be placed on the lower performing student. To maintain our competitive economic edge, educational policies such as NCLB should change their accountability systems to not just focus on bringing up the lowest performers but on extending and raising the highest achievers (Xiang et al., 2011).

It should be noted that NCLB is currently in the process of reauthorization and states are able to apply for waivers, which may effectively eliminate the specific verbiage of AYP (Kress, Zechmann, & Schmitten, 2011). However, removing the accountability for the academic performance of each subgroup, which includes students of differing Ethnicity and socioeconomic status, is not being considered. By forcing schools and states to report out on the performance of each subgroup, NCLB has exposed the disparate academic achievement levels of our nation's students (Chambers, 2009; Giroux & Schmidt, 2004; Kress et al., 2011).

Statement of the problem

Public schools are examining their policies and instructional practices to address the achievement gap exposed by the reporting requirements of NCLB (Wenglinski, 2004). As accountability measures and stakes rise, there is a call for an improved use of scientific evidence to inform educational policymaking (Wiseman, 2010). In terms of the achievement gap, national studies at the secondary level show when students are grouped homogeneously according to ability there is a rise in achievement inequality between the groups (Gamoran & Mare, 1989; Hoffer, 1992). In fact, in the lower ability groups, there is a disproportionate number of minority and economically disadvantaged students (Ansalone 2006, 2009; Gamoran & Mare, 1989). Ability grouping involves separating students into groups according to their perceived academic abilities (Biafora & Ansalone, 2008; Callahan, 2005; Cooper, 1999; Slavin, 1991). This separation can occur within classes or can be a structural adaptation in which students of higher academic ability are placed in classes separate from their lower performing peers (Ansalone & Biafora, 2010; Slavin, 1991). The students' prior academic achievement is usually the determining factor in whether students are placed in the higher performing group or track (Archbald, Glutting, & Qian, 2009; Ballon, 2008; Slavin, 1991). These special classrooms for

the higher functioning (i.e., gifted) students are accompanied by curricula and instructional practices different from the classes containing the lower functioning students. The objective of this difference is to provide a level of education commensurate with the high cognitive levels of gifted students (Ansalone & Biafora, 2010; Biafora & Ansalone, 2008; Preckel, Gotz, & Frenzel, 2010). Indeed, this grouping or tracking of the gifted students has empirical evidence of its benefits for the gifted student and hence, is used to support its practice (Goldring, 1990; Kulik & Kulik, 1982; Rogers, 1993; Rubin, 2003; Shields, 2002). Advocates of this homogeneous ability grouping hold that teachers can best meet the needs of students whose abilities, motivation, and aspirations are similar (Allan, 1991; Benbow & Stanley, 1996; Gamoran, 2009; Oakes & Guiton, 1995).

Alternatively, this practice of homogeneous grouping has not generated nearly the amount of beneficial evidence for students in the lower ability groups. In fact, the practice of ability grouping has shown to depress the academic achievement of students placed in the lower groups (Ansalone, 2000; Carbonaro, 2005; Oakes, 1992; Oakes & Guiton, 1995; Slavin, 1991). National studies at the secondary level reveal an increased gap in the achievement of students in the high and low ability groups (Callahan, 2005; Gamoran, Nystrand, Berends, & LePore, 1995; Hoffer, 1992).

When examining the students generally placed in each of these groups, one finds disproportionality. African-American, Hispanic, and students of low socio-economic status, for example, are more likely to be placed in lower ability groups (Ansalone 2001, 2003; Carbonaro & Gamoran, 2002; Goodlad & Oaks, 1988; Oakes, 1987). As many as 700 studies have explored the nature and consequences of tracking (Ansalone, 2006). Most show that tracking adversely affects the academic achievement and career paths of our disadvantaged students (Ansalone,

2006; Burris & Wellner, 2005; Gamoran & Mare, 1989; Hallinan & Kubitschek, 1999). In a study done by Chambers (2009), ability grouping practices stunted the achievement of students in the lower groups, thereby solidifying and intensifying disparities in performance between the groups. Therefore, the unequal allocation of instruction between these ability groups may result in the widening of the achievement gap between high and low level classes over time (Chambers, 2009; Gamoran & Mare, 1989; Gamoran et al., 1995). Despite the research, ability grouping remains a practice in approximately 80% to 85% of US high schools (Archbald et al., 2009). Therefore, if NCLB is asking educators to report on the progress of each subgroup of students, the practice of grouping our students by ability is problematic in getting all students to meet common standards.

Students are segregated into different classrooms according to past measures of academic performance (Archbald et al., 2009; Ballon, 2008; Oakes & Guiton, 1995). The theoretical framework suggests that grouping students according to ability facilitates the teaching process by making it easier for the teacher (Ansalone, 2009; Keliher, 1931). Other grouping arrangements may enhance learning but make teaching more difficult. Baines, Blatchford and Kutnick (2003) summarize the dilemma; "achieving a strategic balance is vital for effective teaching and learning but is one of the most difficult dilemmas facing teachers" (p. 10). Adding to the problem are findings which show segregating students according to achievement perpetuates the gap in academic achievement along race and class lines (Ansalone, 2000, 2004, 2009, 2010; Argys, Rees & Brewer, 1996; Burris & Wellner, 2005; Chambers, Higgins, & Scheurich, 2009; Gamoran & Mare, 1989; Gamoran et al., 1995; Gamoran & Carbonaro, 2003; Mallery & Mallery, 1999; Mickelson, 2001; Oakes, 1992).

The existing literature on the practice of tracking and/or ability grouping in our schools is ubiquitous. Despite the research, many parents and educators believe ability grouping benefits high achievers, and therefore, an entrenched culture of ability grouping remains firmly in place throughout America's high schools (Ansalone & Biafora, 2010; Biafora & Ansalone, 2008; Burris & Wellner, 2005; Preckel et al., 2010). These parents fear that if their honors students are placed with non-honors peers, they will be exposed to a less challenging curriculum, which will lower the standards for their children. Proponents of ability grouping, including the parents of honors students, want the honors students in ability groups (separate from the other students) so that they can reach their full potential without being hindered by the lower achieving students (Ansalone, 2010; Burris & Welner, 2005). Opponents of tracking, site research that shows the race, ethnic, and SES disproportionality of students found in lower track (i.e., regular) courses vs. higher track (i.e., honors) courses. This disproportionality contributes to the achievement gap (Ansalone, 2006; Burris & Wellner, 2005; Livingston, 2010; Venzant, 2006). Furthermore, this arrangement hinders students who are in the lower track courses from meeting common standards and schools from meeting NCLB requirements.

Theoretical/conceptual framework

The theoretical framework of this study is: Instruction can be targeted more efficiently when students are homogeneously grouped (Allan, 1991; Barnard, as cited by Tyack, 1974; Benbow & Stanley, 1996; Gamoran, 2009; Gamoran & Weinstein, 1998; Keliher, 1931; Oakes & Guiton, 1995; Turney, 1931). In his support of the ruling *Parents Involved v. Seattle*, Supreme Court Justice Clarence Thomas referred to this framework writing, "schools frequently group students according to ability as an aid to efficient instruction" (Chayte, 2009–2010, p. 630). Additionally, Hallinan and Sorensen (1983) referred to homogeneous grouping as the best

way to manage students and keep them attentive; and Oakes (1987) stated, "tracking...was adopted as the means for managing student diversity" (p. 129). Perhaps Slavin (1987) articulated the theoretical framework for ability grouping best when he said it is:

> supposed to increase student achievement primarily by reducing the heterogeneity of the class or instructional group, making it more possible for the teacher to increase the pace and level of instruction for high achievers and provide more individual attention, repetition, and review for low achievers. It is supposed to provide a spur to high achievers by making them work harder ... and to foster success within the group of low achievers, who are protected from having to compete with more able age mates. (p. 296)

Oakes and Guiton (1995) also noted that schools fit the social order and use educational structures to match students and courses to accommodate individual differences and further societal goals. The educational structure adopted most often is ability grouping and the means for such grouping views prior academic achievement as the most significant variable for group selection (Archbald, Glutting, & Qian, 2009; Ballon, 2008; Mickelson, 2001).

Purpose of the study

The purpose of this study is to determine if there is a tangible, measurable academic benefit to homogeneously grouping high school honors English students in a diverse, suburban school district in Washington State. Specifically, the effect of the independent/predictor variable—type of grouping (mixed-ability/like-ability)—on the dependent/outcome variable of student achievement will be analyzed when controlling for student variables associated with student achievement. A standardized 8th grade state assessment in reading will serve as a pre-treatment variable and the standardized 10th grade state assessment in reading will serve as a post-treatment variable. In addition, the PSAT critical reading scores of students will be used as

post-treatment measures. By concentrating on the variable of student grouping and analyzing quantitative data collected both before and after student exposure to the independent variable, the objective of this study is to produce research-based evidence to assist policymakers, educators, and parents in their decisions on whether to group students by ability. Furthermore, a goal is to have data that assist in the development of structures that will maximize the learning and achievement of all students.

Research questions

The present study is couched in the aforementioned conceptual framework and is guided by the following overarching question:

- To what extent, if any, does ability grouping of high achieving students defined as 9th and 10th grade honors English students in a suburban Washington State school district, affect their performance on state and pre-college assessments of reading and writing achievement when controlling for student mutable variables?

The present study is guided by the following subsidiary research questions:

- To what extent, if any, does placement of 2009–2011 honors English students in homogeneous ability grouped English classes have on their subsequent performance on state and pre-college assessments of reading and writing achievement?
- To what extent, if any, does placement of 2010–2012 honors English students in homogeneous ability grouped English classes have on their subsequent performance on state and pre-college assessments of reading and writing achievement?
- To what extent, if any, does placement of 2009–2011 honors English students in heterogeneous ability groups have on their subsequent performance on state and pre-college assessments of reading and writing achievement?

- To what extent, if any, does placement of 2010–2012 honors English students in heterogeneous ability groups have on their subsequent performance on state and pre-college assessments of reading and writing achievement?

Significance of the study

Jefferson advocated for public education as a means for maintaining democracy, "If a nation expects to be ignorant and free in a state of civilization, it expects what never was and never will be" (Tanner & Tanner, 2007). Mann and Dewey championed this progressivist notion by advocating that a major function of public schools was to unify students of diverse backgrounds. A contradictory philosophy sees schools as a meritocracy and views students being served and rewarded according to the merits of their work (Lemann, 1999). This capitalistic idea of schooling places schools in a context of competition (Powell, 2001). Couple the concept of meritocracy with Freidman's (2005) reference to the shrinking of the world marketplace, and American schools are now routinely compared to, and seen as competing with, schools in other countries (Cavanagh, 2012). The flattening of the world economically and the necessity to compete internationally have expedited the creation of common standards or a more essentialist path for American schools (Thurlow, 2012). The adoption of common standards is seen as necessary in order to facilitate future U.S. competitiveness and to maximize potential profitability in a global market.

Whether one has a progressivist or essentialist bent, demands of change in American schools is universal. The Common Core State Standards have been adopted in 45 states (Common Core State Standards, n.d.). All students leaving school "college and career ready" is an emphasis of the Common Core State Standards (Common Core State Standards, n.d.; Thurlow, 2012). How can American schools accomplish the goal of getting all students to meet

common standards? Some advocate homogeneous ability grouping so students can reach their maximum potential without being hindered by lower achieving students (Ansalone, 2009; Burris & Wellner, 2005). Others site the race, ethnic, and SES disproportionality of students found in lower ability groups and consider this a major contributor to these students not achieving at the same levels as their separately grouped peers (Ansalone, 2006; Burris & Wellner, 2005; Livingston, 2010; Venzant, 2006). Demonstrating an empirically verifiable benefit to the upper-end students is the major reason to ability group homogeneously since there is overwhelming evidence to support that lower performing students' academic growth is hindered by their placement separate from higher performing honors students (Archbald & Keleher, 2008; Hallinan & Kubitschek, 1999; Rubin, 2003; Rubin & Noguera, 2004). If there is no empirical evidence showing a benefit to homogeneous ability grouping of high school honors English students, why continue the practice?

The answer to this question may help educational practitioners with a dilemma that has faced them for over a century. If the answers to the research questions find no significantly positive benefit to grouping honors English students homogeneously then past and future studies can be used to explore what structural, technical, pedagogical, and socio-cultural components are necessary to achieve an outcome where both higher and lower performing students show maximum academic benefit. If further studies have similar findings then the practice of homogenous ability grouping across classrooms needs serious reconsideration. A discontinuation of homogeneous ability grouping would have an enormous effect on the way high schools are structured and the means by which students are taught. If a benefit to homogenously grouping honors English students is found in the study then further studies are needed concerning how to raise the academic achievement of lower performing students without

bringing down higher performing students. Educators, parents, and students do not want to sacrifice excellence for equity; they want them both (Allan, 1991; Carbonaro, 2005; Gallagher, 1997).

Slavin (1995) classified studies on tracking as falling in one of two types: 1) high track/low track or 2) track/no track. Slavin (1995) found high track/low track studies problematic. These studies, he says are like "comparing apples to oranges" (p. 221). Furthermore, he states that track/no track studies "are far more meaningful" (p. 221). This study follows the advice of Slavin (1995) and is a track/no track comparative study.

Limitations

There are several limitations relative to this study. Classroom ability grouping will be the focus of this research study. While within-class ability grouping is a component of the heterogeneous classroom, the scope of effective or quality within-class ability grouping is beyond this study. The study was conducted in one suburban school district in Washington State. The sample size, population of students, quality of instruction, delivered curriculum, the sensitivity of the assessments, school variables, and optional nature of the PSAT as a dependent variable are also limitations.

Delimitations

The study is delimited to only students in ninth and tenth grade English classes who take the grade 8 and grade 10 state assessments in reading and the PSAT in grade 10 and/or 11. Furthermore, the study is delimited to honors English students who attend one of two high schools in a suburban, Washington State school district with four comprehensive high schools. The high schools studied are in a Washington State suburban district each with similar size (i.e., 1800–2000 students) and with similar demographics.

Definition of terms

In the high number of research studies done on the topic of tracking, generally, three types of tracking are identified: comprehensive full-day grouping, within-class ability grouping, and regrouping for specific subject areas—mixed ability or like-ability groups (Allan, 1991; Ansalone, 2009; Mosteller, Light & Sachs, 1996). This study is focused on regrouping for specific subject areas, specifically 9th and 10th grade English. The two types of groupings are mixed-ability (i.e., heterogeneous) and like-ability (i.e., homogeneous) groupings. The following terms are used in this dissertation:

- *Ability grouping*: practices that assign students to different educational environments based on past academic achievement (Schofield, 2010), many times used synonymously with tracking.
- *Curriculum differentiation*: providing students with classes at different levels of difficulty. In U.S. secondary schools, students are often allowed the choice of *basic*, *regular* or core, and *advanced* or honors. In its stronger form, this is referred to as tracking (Schofield, 2010).
- *Detracking* (or untracking): placing students in heterogeneous ability groupings or mixed-ability groups. (Allan, 1991; Ansalone, 2006, 2009; Gamoran, 1989, 1993; Mosteller et al, 1996; Patton, 2010).
- *Gifted:* A student with unusually high prior achievement scores. Often perceived as having higher cognitive ability, synonymous with "honors" (Goldring, 1990).
- *Heterogeneous grouping*: synonymous with mixed-ability grouping.
- *Homogeneous grouping:* synonymous with like-ability grouping.
- *Like-ability grouping*: placing students of similar abilities into the same class or group.

- *Mixed-ability grouping:* placing students of differing abilities into the same class or group.
- *Setting*: The term used predominantly in England and countries other than the U.S. to describe ability grouping for one or two classes (Ansalone, 2003; Schofield, 2010).
- *Streaming*: The term used predominantly in England and countries other than the U.S. to describe tracking (Ansalone, 2003; Schofield, 2010).
- *Tracking*: separating students into homogeneous or like-ability groups (Allan, 1991; Ansalone, 2006, 2009; Gamoran, 1989, 1993; Mosteller et al., 1996; Oakes, 1985; Patton, 2010).

CHAPTER II

LITERATURE REVIEW

Introduction

It remains unclear if and how placement of students into mixed-ability groups affects the achievement of the gifted student. What is clear is that research shows there is disproportionality in the race, ethnic, and SES of students found in lower track courses vs. higher track courses. This disproportionality contributes to the achievement gap (Ansalone, 2006; Burris & Wellner, 2005; Livingston, 2010; Venzant, 2006). Furthermore, this arrangement hinders students who are in the lower track courses from meeting common standards and schools from meeting NCLB requirements (Bloomfield & Cooper, 2003; Day-Vines & Patton, 2003; Sunderman, 2003).

The purpose of this study is to determine if there is a tangible, measurable academic benefit to homogeneously grouping high school honors English students. Specifically, an analysis of the effect of the independent variable—type of grouping (mixed-ability/like-ability)—on the dependent variable of student achievement. A standardized 8^{th} grade state assessment in reading will serve as a pre-test and the standardized 10^{th} grade state assessment in reading will serve as a post-test. In addition, the PSAT Reading scores of students will be used as an additional post-test measure. By concentrating on the variable of student grouping and analyzing quantitative data collected before, during, and after student exposure to the ability grouping independent variable, the objective of this study is to produce research-based evidence to assist policymakers, educators, and parents in their decisions on whether to group students by ability. Furthermore, a goal is to have data that will assist developing structures that will maximize the learning and achievement for all students.

The review of the literature is divided into the following sections: 1) historical development of tracking/ability grouping, 2) worldwide and national standards movement, 3) conceptual/theoretical framework, 4) achievement gap, 5) tracking, 6) detracking, and 7) synthesis of research.

Literature search procedures

Literature for review was found using JStor, ERIC, Proquest, and Google Scholar. Literature was acquired both electronically and through hard copies found in university libraries. Search terms used included: *ability grouping, achievement gap, detracking, gifted education, heterogeneous grouping, high school English, homogeneous grouping, mixed ability grouping, tracking, secondary English, setting,* and *streaming.* These terms were used in isolation or combination to produce search results.

Limitations of review

Books were used for historical and theoretical background information only. Journal articles involving secondary schools were the overwhelming focus, although several studies relating to elementary levels were included to enrich the findings regarding the effects of grouping on student achievement. Studies that met the following criteria were included: reported effect sizes of greater than 0.25 (if reported), experimental, quasi-experimental, non-experimental with control groups, causal-comparative, and qualitative; were peer-reviewed or government reports; reported at least statistical significance ($p \leq .05$); were published within the last 30 years unless considered a seminal work.

Criteria for inclusion

Primary and secondary sources as well as both periodicals and dissertations were included in the literature review. Books were only included for establishing a historical or

theoretical background or if they were considered seminal works in their fields. Articles in languages other than English were not included.

Historical perspective of tracking

Formalized tracking in schools can be traced back to as early as 1800 when teachers were few in number. Free appropriate public education was becoming increasingly difficult due to large class sizes resulting from a lack of trained teachers and a lack of funds (Keliher, 1931). To address this problem a monitorial plan of instruction was adopted. In this plan, the teachers trained the brighter, older students (i.e., monitors) who then taught groups of the younger students. This system allowed the monitorial schools of Philadelphia in 1819 to have one teacher teach over two hundred pupils. This monitorial system began the shift from the one-room schoolhouse with children of multiple ages and abilities to multi-room schools that grouped and organized students differently (Keliher, 1931). Educational theorists such as Horace Mann led the call for schools to "replace the heterogeneous grouping of students with a systematic plan of gradation" (Tyack, 1974, p. 44). John Philbrick convinced the Boston school board to adopt a model for this type of grouping. Known as the *egg crate* School, the Quincy School was established in 1848; Philbrick became principal (Tyack, 1974). This school was built so that each teacher had his or her own classroom for the one grade level of student that he or she taught. As recorded by Tyack (1974), Philbrick believed that scholars should be "divided according to their tested proficiency" (p. 45). By 1860, most of the schools in America separated their students into grades (Tyack, 1974).

The concept of grouping students according to ability continued throughout the nineteenth century. In 1886, W. J. Shearer created a plan known as the Elizabeth plan. The essential feature was to divide grades into sections so that "pupils could be grouped together by

attainments" (Keliher, 1931, p. 21). Additional plans were adopted each with their objective to classify students according to ability and then to have the classes move forward at rates appropriate to their abilities. Further supporting the efforts of educational practitioners and theorists in the grouping of students was the adoption of compulsory attendance and child labor laws between 1880 and 1900 (Keliher, 1931). The effect of these laws was to increase the number of children attending schools while compelling students who otherwise may have quit to remain there. The most widespread device for addressing this problem of individual differences was homogeneous grouping of students (Keliher, 1931).

In the post-Civil War period, large numbers of farm workers and blacks, in pursuit of a better life, migrated and settled in the north. Due to the increased differences in student backgrounds and abilities, tracking increasingly became accepted practice (Ansalone & Biafora, 2008). In the 1850 decision *Roberts v. City of Boston,* the Massachusetts Supreme Court rejected an equal protection challenge brought by black parents (Ansalone, 2010; Findley, 1989). Chief Justice Shaw articulated a theory that has endured throughout subsequent litigation. He maintained that schools had the "plenary authority to arrange, classify, and distribute students as they think best adapted to their general proficiency and welfare" (*Roberts v. City of Boston*, 1850). Boston's practice of a dual-track school system, therefore, was upheld as a necessary practice to arrange, classify, and place students to satisfy their general proficiency and welfare (Findley, 1989). The U.S. Supreme Court in *Plessy v. Ferguson* (1896) affirmed this "separate but equal" doctrine and allowed the segregation of our schools along racial grounds.

Heading into the twentieth century, the American populace continued to change. Adding to the numbers of blacks who were moving to the north, were the immigrants who were making America their new home. Ayres (1909) made the case that immigrant children were retarded and

schools were wasting resources in their attempt to provide immigrants an education equivalent to that of American youth. To deal with this influx of newcomers, the one-room schoolhouse needed to be replaced by more factory-like structures (Ansalone, 2004; Tyack, 1974).

During this early twentieth century period, Alfred Binet developed an IQ test as a scientifically valid measure of a person's intelligence. He contended that this test could be used to determine a student's appropriate placement in our schools (Tyack, 1974). Goddard (1914), Terman (1916), and Yerkes (1915) followed the work of Binet, each of whom extended the concept that intelligence is inherited and stable. In fact, Terman (1916) bluntly referred to Spanish-Indian, Mexican, and Negro people as dull. He stated, "children of this group should be segregated in special classes and be given instruction which is concrete and practical" (p. 92). Therefore, the educational practice at the turn of the twentieth century became a way to prepare students for their appropriate place in the workforce. High ability students were given access to advanced academic training and students who lacked academic acumen were placed in lower level tracks and trained for vocational positions (Ansalone, 2006: Cooper, 1996). Although students had advanced through school at different rates based on their performance, this period reinforced the practice of placing students according to ability (Findley, 1989).

Educational theorists led by John Dewey extended the work of the NEA's Committee of Ten and began to erode the concept of tracking (Tanner & Tanner, 2007; Tyack, 1974). They cited the importance of heterogeneous grouping in public schools as a unifying practice to bring together students of diverse backgrounds. Dewey advanced the idea that knowledge should be equally available to all members of society, not just the privileged. The Committee of Ten declared "the rigorous training of the mind through academic subjects would best fit anyone for the duties of life" (Tyack, 1974; p. 58). These ideals of an inclusive education were extended by

the seminal research findings of the *Cardinal Principles* (1918) and Ralph Tyler's 1930s *Eight Year Study* (Tanner & Tanner, 2007). Due to the admonition of Dewey and others, and a decline in immigration due to WWI, ability grouping began to wane between the mid-1930s and the 1950s. The onset of the civil rights movement and the beginning of President Johnson's war on poverty rendered the genetic rationale for intelligence less acceptable (Oakes et al., 1997).

The outbreak of World War II in the 1940s and later the Cold War stymied the momentum of mixed ability or heterogeneous grouping. These two monumental events launched the American education system into a more discipline-centered focus (Chayte, 2010; Tanner & Tanner, 2007). Indeed, James Conant's influential report The *American High School Today* published in 1959 advocated a return to the practice of grouping students according to ability. Conant (1959) asked the question: Can a comprehensive high school educate "those with a talent for handling advanced academic subjects" (p. 15)? Conant generally found that the academically talented student was not being sufficiently challenged. Conant concluded that students of similar ability should have almost identical programs. "In the required subjects...student should be grouped according to ability, subject by subject" (p. 49).

In addition, in this era of the space race and the national demand to out-compete the Soviets, Bruner issued *The Process of Education* (1960) report. Bruner advanced the notion that in the long-range crisis in national security a successful resolution will depend on a well-educated citizenry. Bruner stated "excellence must not be limited to the gifted student; but the idea that teaching should be aimed at the average student in order to provide something for everybody is an equally inadequate formula" (p. 70). Although not specifically advocating ability grouping, his work promoted an educational focus on the academically gifted.

These shifts in the American educational system were occurring in a time of great social unrest and change. In 1954, the United States Supreme Court in their landmark ruling *Brown v. Topeka Board of Education* held that segregated education was inherently unequal. The Cold War, together with the desegregation demands of the 1950s, prompted a movement to separate gifted students in order to provide them a special education. Southern states used ability grouping to avoid desegregation orders, and the northern cities used ability grouping in response to the large migration of blacks (Ansalone, 2006; Chayte, 2010). Furthermore, in 1964, the United States Congress, by enacting the Civil Rights Act of 1964, codified the court's finding (CRA, P.L. 88-352). This act provided many protections, among them, prohibiting racial segregation in our public schools. Additionally, in 1965, Congress authorized the Elementary and Secondary Education Act (ESEA) in support of Lyndon Johnson's war on poverty. President Johnson emphasized the importance of all children having access to a quality education as a means to leading productive lives. In 1975, Congress passed the Education for All Handicapped Children Act (EHA, P.L. 94-142). This act required federally funded schools to provide a free appropriate public education (FAPE) to students with disabilities. Individually and collectively, these pieces of legislation and the findings of the court had a lasting and far-reaching effect on United States policy and practice. In terms of educational effects, public schools were required to desegregate racially and to include all children regardless of the presence of disabilities. Despite these key legislative and judicial findings, tracking remains a practice in not only American schools but also schools worldwide.

The standards movement in American education

In 1983, the National Commission on Excellence in Education released their report; *A Nation at Risk: The Imperative for Educational Reform.* This report found that many of our

students were functionally illiterate, SAT scores were trending downward, and students were in need of an increasing number of remedial courses in college. The report concluded that these findings and trends showed that our educational system was threatening our nation's future. In the end, the report served to heighten and solidify federal involvement in public education (Hewitt, 2008). *Goals 2000*, which became law in 1994, furthered the federal government's role by providing funding to states that were implementing school reforms and developing education standards and assessments. Signed into law in 2002, the 2001 No Child Left Behind Act substantially increased the testing requirements and set demanding accountability standards for schools, districts, and states (Bloomfield, 2003). Concerns gradually arose that state standards varied widely. Comparing student performance on state assessments with their performance on the NAEP clearly demonstrated variability in the rigor of each state's tests (Rothman, 2012). Due to this variability, content standards are becoming nationalized and are currently referred to as the Common Core State Standards (Rothman, 2012). To assess a school's progress in meeting these common standards, public school districts and their schools are being asked to demonstrate that they are making progress with students in each of the applicable student subgroups. NCLB does not use a broad measure of a school's overall averages but requires schools to report on the achievement of a number of student subgroups to determine school effectiveness (Fusarelli, 2004). NCLB has as one of its goals to diminish the achievement gap between minority and nonminority children, especially between low socio-economic students and their more affluent classmates (Bloomfield & Cooper, 2003; Day-Vines & Patton, 2003; Sunderman, 2003). It should be noted that NCLB is currently in the process of reauthorization and states are applying for waivers, which would eliminate the penalties associated with AYP (Kress et al., 2011). However, removing the accountability for the academic performance of each subgroup, which

includes students of differing Ethnicity and socioeconomic status, is not currently being considered. By forcing schools and states to report on the performance of each subgroup, NCLB has exposed the disparate academic achievement levels of our nation's students (Chambers, 2009; Giroux & Schmidt, 2004; Kress et al. 2011).

With the advent of standards-based school reform and the resultant school accountability measures, attention on school effectiveness has gone beyond general measures of effectiveness to a focus on the academic achievement of all students. In large part, due to the reauthorization of ESEA and the signing into law of No Child Left Behind (NCLB) in 2002, schools are now required to report on the academic achievement of a number of student subgroups. As part of this reporting requirement the issue of the gap between student subgroups (African-American, Hispanic, Native-American, English Language Learners, special education, and low-income students) compared to their Asian and white peers has become glaringly apparent.

Curriculum decisions in most American high schools are largely centered on mechanisms for placing students of different academic abilities into classes at the appropriate level (Oakes, Selvin, Karoly & Guiton, 1992). This matching of students to different high school programs has carried with it racial, ethnic, and social-class implications. ELL, poor, and minority students are more often placed into low-level academic tracks where middle and upper class whites are placed into the higher academic tracks (Archbald, et al., 2009; Ballon, 2008; Oakes, 1995; Oakes et al., 1992). This disproportional representation of students in lower track levels as compared to students in the higher track levels greatly concerns educators interested in closing the achievement gap (Burris & Wellner, 2005). Ansalone (2006) equates to this disproportionality of blacks in the lower tracks as equivalent to returning to the days of Jim Crow schools, where

blacks were provided an education with "crushing limitations with little or no opportunity to learn" (p. 146).

Brown vs. Board of Education (1954) mandated that states providing an education to their people must provide that education to all on an equal basis. The implementation of school ability grouping or tracking undermines the doctrine of *Brown* when students of color are assigned to lower track classrooms (Venzant, 2006). Because low achievers are more likely to be assigned to the lower tracks, tracking reinforces the initial differences between the student groups, therefore, widening the gap in achievement (Chambers, 2009; Gamoran & Mare, 1989). Although desegregation was mandated by the *Brown* decision, a subtle form of racial segregation has continued within our schools in the form of tracking (Archbald, et al., 2009; Mickelson, 2001). The association of tracking with perpetuating the achievement gap is widely reported in educational literature (Ansalone, 2006; Burris & Wellner, 2005; Gamoran & Mare, 1989; Gamoran, et al., 1995; Hoffer, 1992; Mickelson, 2001). Students who are African-American or Latino, and students from low-income families, are routinely placed into tracks where they are exposed to different curricula with low levels of rigor compared to their Asian American, White, and higher socio-economic peers (Mickelson, 2001; Oakes, et al., 1992; Venzant, 2006). Attewell and Domina (2008) studied whether "there is good evidence that upgrading the content of high school courses improves student performance on indicators such as test scores" (p. 51). Taking a more rigorous curriculum was associated with higher test scores (Attewell & Domina, 2008). They found, however, there were significant discrepancies in access to challenging courses that could not be explained by prior academic performance alone. The disparity in student access was along family SES lines more so than ethnicity. Students along race and economic lines are not exposed to the same rigorous learning opportunities as their higher SES

peers (Attewell & Domina, 2008). This disproportionate placement and access to a more intense curriculum contributes to the lower levels of achievement among the students identified by the subgroups used in NCLB.

The idea of framing the disparity in student academic achievement as an *achievement gap* comes under fire from Chambers (2009). In her study, Chambers argues instead of focusing on the outputs (student standardized test scores) educators should focus on the educational services students of color receive because of their placement in differing educational tracks. She offers the term *receivement gap* to encourage educators to examine the practices that occur which facilitate the disparity in academic outputs. This is a nuanced but important distinction. Instead of just focusing on the outputs of tracking, educators must place even more emphasis on examining why students in different levels continue to move apart on measures of academic achievement. Callahan (2005), in her study on the effects of tracking on English Language Learners, found that these students placed in the lower track classes are exposed to less rigorous content and fewer learning opportunities than students in high track placements did. Callahan (2005), like Chambers (2009), advocates for a renewed look toward content-area instruction rather than the attributes of the learners. Similarly, researchers have found that track placement accounts for variation in both student growth and achievement primarily because of the difference in content-area coverage (Carbonaro & Gamoran, 2002; Hallinan & Kubitschek, 1999). Many have found that students in the lower tracks are exposed to topics and skills that were less demanding, while those in the upper tracks were taught more complex thinking and problem solving tasks (Ansalone, 2004; Carbonaro & Gamoran, 2002). This interest in the differences of the inputs occurring within the classrooms between the tracks helped reenergize the study of within-classroom instructional strategies (Tomlinson, 2006). That is, if separating

students into different tracks allows the high-performing students to move ahead while leaving the lower-performing students farther behind; what can be done within the classroom to raise the achievement of all of the students?

Theoretical/conceptual framework

Before discussing the central theoretical framework of the practice of tracking, it is important to recognize that even the work of theorists in other fields has been used to support the practice. The economic theorist Vilfredo Pareto has been cited by theorists and practitioners in many academic fields including education (Argys et al., 1996). Efficiency—economic or otherwise—is realized by the recognition that human society is divided into the *ruling class* and *the masses*. Pareto (1902) identified the cause for this separation as a differential distribution of talent, skill, and intelligence among men (Barkley, 1955). Paretian thinking contends it is sound to realize that the academic elite deserve special accommodations for they will produce the greatest societal rewards (Barkley, 1955). Since school resources are limited in terms of material and human resources, supporters of this perspective view the perceived cognitive achievement gains of the high ability student as the most significant and valuable contribution of tracking (Ansalone, 2010; Kulik and Kulik, 1982; Mosteller et al., 1996). Similarly, Bandura (1997) contended that social learning theory supports the ability grouping of students. According to social learning theory, students will achieve at levels similar to which they are associated. Therefore, by grouping students according to ability, a school offers the high achievers a relatively high academic benefit (Epple, Newlon, & Romero, 2002).

As cited in the historical background section of this study, the most widespread, endemic reason for homogeneous grouping is to allow educators to most efficiently meet the individual or different learning needs of students (Ansalone, 2009: Burr, 1931; Keliher, 1931; Tyack, 1974).

The monitorial system described earlier was introduced as an attempt for the educational system to meet the needs of more children in the schools efficiently (Keliher, 1931). The concept of ability grouping's increased efficiency is widespread in literature. According to Turney (1931) "the aim of ability grouping is to bring together pupils who will be able to work together and to progress together under conditions permitting the fullest possible development of the individuals involved" (p. 22). Current Supreme Court Justice Clarence Thomas in his support of the ruling *Parents Involved v. Seattle* (2007) referred to the theoretical framework when he wrote, "schools frequently group students according to ability as an aid to efficient instruction" (Chayte, 2009–2010, p. 630). The notion of efficiency also is advanced by social theorists who believe that tracking increases efficiency by contributing to the proper selection and routing of national human resources (Epple et al., 2002). That is, with the increasingly diverse nature of America's schools and with limited financial resources; an efficient distribution of resources—including human—is necessary to achieve the goals our nation has for its schools.

Turney (1931) theorized four reasons for ability grouping. It can:

- facilitate instruction by allowing for individualization;
- empower instructors to adjust their teaching techniques to match the ability level of the students;
- reduce the boredom of advanced students due to the separation of slower students; and
- encourage participation of the lower ability students since they will not have to compete with their more capable peers.

Similarly, Keliher (1931) stated that the objective of ability grouping is to have the "different classes...go forward at rates appropriate to their varying abilities" (p. 22). Keliher further delineated the theory behind homogeneous grouping by identifying the assumption

"homogeneity of grouping tends to bring superior learning results" (p. 91). Therefore, the teaching of students grouped according to ability is seen by many as necessary for achieving the best learning outcomes for both the lower and higher ability group.

Teachers rely on this theory of instructional efficiency because they have found that creating lessons to meet the needs of learners with varying needs is difficult. Argys et al. (1996) believed teachers see tracking "as a way to reduce the range of performance and motivation...making teaching easier and preventing less able students from 'holding back' those with greater academic talent" (pp. 624-625). What often happens is that while one student may find the level of instruction satisfactory, another tends to require more time. Hallinan and Sorensen (1983) suggested that low ability students benefit from this segregation because the teachers can provide them with the appropriate curriculum and pace of instruction. Kerckhoff (1986) also contended that high ability students can accelerate without having to wait for their less competent peers.

Because of the differences in learning rates, in heterogeneous classes the fast learner may become bored by the lengthy and simple explanations provided to slower learners. When the needs of the slower learners and the more advanced learners are not met and if the teaching is not geared to the appropriate ability level of the students, problems occur (Khazaeenezhad, Barati, & Jafarzade, 2012). Consequently, boredom may lead to classroom management issues. Ability grouping is thought to be a panacea for classroom management problems (Ireson, Hallam, Hack, Clark, & Plewis, 2002). It is believed that students grouped by ability are easier to manage and keep attentive (Hallinan & Sorensen, 1983). Teachers report a need to educate students in ability groups as a classroom management tool in order to work more effectively with the disparate range of academic needs within classes (Ansalone & Biafora, 2004). Teachers see ability

grouping not necessarily as a benefit to one student group or another but rather as a structure to make classroom life more manageable.

As mentioned earlier, Conant (1959) concluded that in the required courses students should be grouped according to ability. According to Bruner (1960), excellence is not just for the gifted student but also for all students. The challenge with grouping students heterogeneously is being able to meet the needs of all learners. One strategy used in heterogeneous grouping is for the teacher to teach to the middle. Conant and Bruner argued that this strategy is not a desirable alternative. Therefore, concomitant with the social and political pressures of the times, they facilitated the reestablishment of a theoretical framework in American education that grouping students according to ability benefits students' learning (Tanner & Tanner, 2007). In the context of the Cold War, the conceptual framework of targeting instruction to students in homogeneous groups is more efficient and is beneficial to all was reestablished (Allan, 1991; Benbow & Stanley, 1996; Gamoran, 2009; Gamoran & Weinstein, 1998; Oakes & Guiton, 1995). Current advocates of homogeneously grouping students hold that teachers can best meet the needs of all students whose abilities, motivation, and aspirations are similar (Allan, 1991; Benbow & Stanley, 1996; Gamoran, 2009; Oakes & Guiton, 1995).

The theory behind tracking is that placing students in homogeneous groups is more effective in allowing teachers to provide students targeted, effective instruction. This instruction, tailored to meet the skill and ability levels of the students, will then allow and promote an optimal level of student academic growth (Archbald, et al., 2009, Callahan, 2005).

Deciding which students are placed into which homogenous group has been an enduring topic of considerable debate in the educational community. Research is clear that this placement decision can have a significant effect on the levels at which students in each group achieve

(Archbald, et al., 2009; Ballon, 2008). Add to the debate the racial disproportionality found between the ability groups and you have an educational, social, and political hot button issue (Ballon, 2008; Oakes & Guiton, 1995; Yonezawa, Wells & Serna, 2002). Some in education may advocate for a tracking system where students are given the freedom of choice. In theory, this freedom of choice would mitigate the issue of disproportionality. Yonenezawa et al., (2002) in their study of ten secondary schools found that "using choice as a tracking tool is unlikely to change the racial and social stratification of track systems" (p. 59). Chambers (2009), in a case study of a high school in a major Midwestern metropolitan area, confirmed these findings.

Archbald, et al., (2009) raised an important question, "If high school tracks did not exist, then there would be no debates over the equity and outcomes of tracking placement decisions" (p. 78). Instead of focusing attention on track placement criteria, it may be more productive to promote student achievement for all students by focusing on instruction. By differentiating instruction, teachers may be able to give all students access to a high-level curriculum (Tieso, 2003; Tomlinson, 2006). One strategy that can be used in schools is to allow for the heterogeneous grouping of students. Cohen and Lotan (1995) conducted extensive research on the sociology of the heterogeneous classroom. They explored academically heterogeneous small groups in elementary classrooms and found that it is possible to produce significant gains in achievement and participation of the low-status students without depressing the participation of the high-status students. Similarly, Cheng, Lam, and Chan (2008) found that group heterogeneity was not a determining factor in students' learning efficacy. They found that group processes of high quality were effective for both low and high achievers. Haberman (1991) believed in the efficacy of heterogeneous grouping when he stated "students benefit from exposure to cultural as well as intellectual heterogeneity" (p. 294). This heterogeneity allows

teachers to use divergent questioning, multiple assignments, and activities, which allow for alternative solutions and responses. The variety found in heterogeneous classrooms fosters an environment where the students can learn from one another (Haberman, 1991).

Efforts to detrack schools to allow heterogeneous ability grouping often have been short lived. Differentiated instruction is a sound instructional practice designed to meet the needs of all students (DiMartino & Miles, 2005; Haberman, 1991; Tieso, 2003; Tomlinson, 2006). A challenge for detracking advocates has been to overcome the resistance to their efforts for a long enough time for detracking to show positive results (Ansalone, 2009; Cooper, 1996, 1999; Oakes & Wells, 1998; Yonezawa & Jones, 2006). VanTassel-Baska et al. (2008) found that differentiated instruction takes two to three years to show effectiveness and to positively affect teacher beliefs in student learning benefits. Hattie (2002) completed a study where he arrived at the powerful conclusion "whether a school tracks by ability or not...appears less consequential than whether it attends to the nature and quality of instruction in the classroom" and, "The learning environments within the classroom, and the mechanisms and processes of learning that they foster, are by far the more powerful" (p. 449). Although the scope of this study doesn't allow for a thorough examination of classroom instructional practices, it is important to establish a premise that if the findings of this study do not show an appreciable difference in student achievement as a result of ability grouping then further study needs to explore the characteristics of a detracked or mixed-ability classroom where high performing students continue to function at a high level.

Tracking

For the purpose of this study, tracking will be used synonymously with homogeneous ability grouping. That is, dividing students into class-sized groups based on a student's

perceived ability or prior achievement (Biafora & Anasalone, 2008; Cooper, 1999; DiMartino & Miles, 2005; Rubin, 2006). This study will also differentiate the various tracks as either honors (i.e., high ability) or regular (i.e., combined middle and lower ability) classrooms. Proponents of tracking operate under the assumption that students benefit from lessons and a classroom environment that are aligned with their cognitive abilities, motivation levels, and interests (Argys, et al., 1996). They also believe that students who are identified as gifted or talented academically will achieve at higher levels when placed in classrooms with students of like-ability than if they were placed in a heterogeneous classroom (Goldring, 1990; Shields, 2002).

Both *A Nation at Risk* (1983) and *A Place Called* School (1984) stressed the need for American schools to provide appropriate opportunities for gifted and talented youth (Goldring, 1990). Providing these opportunities means there is a need to group gifted learners for their learning and socialization, along with the need to move them ahead in some form. Powerful academic effects will be produced when gifted children are grouped with like-ability peers and exposed to differentiated learning tasks and expectations (Rogers, 2007). These studies lend credence to the proponents of tracking who don't want to replace excellence for all students with equity for all (Benbow & Stanley, 1996). Due to the number of scholarly reports and meta-analyses of grouping literature, the support of grouping for meeting the academic and socioaffective needs of the intellectually advanced student is so compelling it borders on malpractice for schools not to use homogeneous grouping appropriately (Allan, 1991; Benbow & Stanley, 1996; Feldhusen & Moon, 1992; Kulik & Kulik, 1982).

Homogeneous grouping of students has been the practice since the turn of the twentieth century in America's public schools (Tanner & Tanner, 2007). There are several reasons found in literature that support the continued practice of tracking. One is that teachers prefer teaching

students in like-ability classrooms. In many surveys, the general percentage of teachers who prefer teaching homogeneously grouped student classrooms is upwards of 75% (Biafora & Ansalone, 2008; Kulik & Kulik, 1982). Kulik and Kulik (1982) further elaborated that this preference is due to the teachers' experiences that teaching students in homogeneous groupings is easier due to having to deal with fewer individual differences and being able to focus their instructional delivery to one level. The premise is that students who are high achievers are also highly motivated and less prone to distraction and off-task behavior (Freedman, Delp, & Crawford, 2005). Therefore, having students grouped by common abilities reduces the time the teacher has to spend on managing the classroom.

Another reason is that instruction in high ability classrooms tends to be more likely to involve higher-order cognitive challenges such as problem-solving and critical thinking tasks (Oakes, 1992). In addition, grouping by ability is one of the primary ways to effectively deliver to the high-ability student the required differentiated curriculum (Feldhusen & Moon, 1992; Schofield, 2010; VanTassel-Baska, 2005; VanTassel-Baska et al., 2008). Tasks in these specially grouped classes are associated with low off-task behaviors and higher levels of student participation, which can be said to involve increasing amounts of instructional discourse (Gamoran et al., 1995). Gamoran et al. (1995) clarified that just because students are on task it does not mean they are cognitively challenged. Regardless, discourse among high ability students clearly has the potential to enrich instruction and promote functioning at higher levels of cognition.

Parents too advocate for their high achieving student to be placed in higher functioning homogeneous student groups. These parents fear that their students will not be challenged enough, and hence, will not reach their potential if placed with lower functioning peers (Oakes &

Wells, 1998). In fact, parents of children who have been tracked are among tracking's strongest supporters (Ansalone, 2010; Perrone, Wright, Ksiazak, & Crane, 2010). Principals are often proponents of tracking within their schools because it allows them to offer parents greater choice and flexibility in the education of their children. Tracking also allows them to offer to parents a more academically oriented track, which serves as an incentive for parents to keep their high functioning students in their school instead of removing them to another school (Biafora & Ansalone, 2008; Epple et al., 2002). Epple et al. (2002) found no research where tracking was viewed as an equilibrium model. They further state that if tracking were eliminated, students who qualify for the higher tracks would show negative effects concerning achievement. See Table 1 for a summary of reviewed research that supports tracking.

Many research studies have found why tracking is still in favor in many of our schools today. There is evidence of quantifiable academic benefit for high functioning students who are grouped together (Allan, 1991; Argys et al., 1996; Goldring, 1990; Shields, 2002). There are also research findings that show that the level of instructional engagement and challenge is greater for students in honors classes than their peers in heterogeneously grouped classes (Allan, 1991; Shields, 2002). Honors students placed in heterogeneous regular courses also reported suffering boredom or a lack of challenge (Gallagher, 1997; Yonezawa & Jones, 2006). The most striking results involved parent attitudes towards tracking. Parents—most notably those whose children are in the upper track—were almost unanimous in their support of tracking as an instructional practice (Ansalone, 2010). Many parents and educators assume that tracking benefits high achievers. This is partly due to parents' perceptions that detracking will oversimplify the curriculum and lower the learning standards for their children (Burris & Wellner, 2005; Keller, 2011; Oakes & Guiton, 1995). Studies such as these, especially the

attitudes and perceptions of parents, leave no doubt about why tracking is still a widely implemented educational practice. Until these parents change their view that detracking is not a zero-sum game in which one student's gain is another student's loss there will continue to be obstacles in dismantling a system in which only a handful of students are held to high standards (Wells & Oakes, 1996).

Table 1

Sample of examined studies the results of which support tracking

	Argys et al, 1996	Shields, 2002	Goldring, 1990	Allan, 1991	Gallagher, 1997	Gamoran, 1993	Ansalone, 2010
Grade Level/Subject of Study	8th & 10th grade math	5th and 8th grade students			173—Elem. 411—MS 287—HS students	Ability grouped 8th & 9th grade English classes	151 Elem. sts. 75 MS. sts 55 HS sts.
Sample Size	3405 students				173—Elem. 411—MS 287—HS students	18 schools/92 English classes	180 parents
Design and Methods	Qualitative - Survey	Qualitative	Meta-analysis	Narrative Review	Qualitative—Survey of students	Students (tests and Qs) Teachers (Qs. Intvws, Obs)	Qualitative/ Questionnaire
Findings (Effect Sizes)							
Strengths					Large Sample	Large Sample	Large Sample
Results Trusted or Tentative	Trusted—large sample size.		Tentative. Randomized studies are preferred to matched pairs.			Trusted	Significantly significant findings.

Findings	Moving students from a low track classroom to a heterogeneous classroom leads to a 8.6 % increase in math scores (good for low sts.). Moving students from an above average class to a heterogeneous class lead to a 8.4 % decrease in scores (negative for high sts.).	Teachers in gifted classes required students to engage in longer-term research assignments (high level of cognition). Students in heterogeneous classes reported lower teacher expectation, less time on task, less homework, less teacher feedback. Homogeneous grouping benefits the most able. Students in regular classes do not suffer emotionally or socially.	Gifted students in homogenous classes outperformed their gifted peers in heterogeneous classes. Greatest advantage was in math and science. Smaller benefit for reading and writing. Gifted students in regular classes had more positive attitudes towards peers. Differences in gifted sts. achievement caused by organization or instructional practices?	Gifted students show positive academic outcomes from some form of homogeneous grouping. Most positive results come from classes that are taught by specially trained teachers and a differentiated curriculum.	Gifted students reported a lack of challenge in their heterogeneous classes.	Ability grouping with certain instructional (high exp., oral discourse, greater teacher effort) and structural (no weak/inexp. teachers assigned) can be effective.	60% of parents who were tracked as students themselves reported that the practice had a positive lasting consequence for future life choices. 98.6 % of parents whose children were placed in the upper tracks reported the placement was helpful for their student. Findings in line with Perrone et al., 2010.

Detracking

Detracking, also known as mixed-ability or mixed-skill grouping, involves grouping students heterogeneously where the skill level of the children within each class varies considerably (Cooper, 1999; Mosteller et al., 1996; Rubin, 2003). This arrangement of students is increasingly being considered and implemented due to the research findings that show students placed in homogeneous, high achieving groups learn more than students in homogeneous, low ability groups (Carbonaro, 2005; Hoffer & Gamoran, 1993: Hooper & Hannafin, 1991). Separating students into like-ability groups produces inequalities in student educational

outcomes between groups (Hoffer & Gamoran, 1993; Gamoran, 1993; Mickelson, 2001). These inequalities create differences across tracks in the quality and quantity of instruction (Oakes, 1985: Gamoran, 1989). Add to this, students of color and low socio-economic status are more often tracked into student groups that impinge on the achievement and academic and social opportunities (Ansalone, 2000, 2004, 2006; Burris & Wellner, 2005; Gamoran, 1989, 1993; Gamoran, et al., 1995; Oakes & Wells, 1998). The practice of tracking then widens the gap in achievement between the high and low track students and therefore, between students who are poor, black, or Latino and those who are wealthier, white, or Asian (Ballon, 2008; Gamoran & Carbonaro, 2003; Mallory & Mallory, 1999; Mickelson, 2001: Oakes & Lipton, 1992; Oakes, 1992). Even in schools that are ostensibly desegregated, Whites retain privileged access to greater opportunities to learn (Mickelson, 2001). This fact strongly contributes to further critical examination of the practice of tracking in our schools.

The inequitable treatment of students often begins in the screening and selection process. The screening process involves students being selecting based on prior achievement, effort, and other background factors. Although there are avenues for parental input, African-American and Latino parents have little access to this knowledge (Carbonaro, 2005, Oakes & Guiton, 1995, Mallery & Mallery, 1999).

The difference in student academic achievement between the high and low tracks is also due in part to instructional inequities (Ballon, 2008; Gamoran & Carbonaro, 2003; Hattie, 2002; Hoffer & Gamoran, 1993). Teachers who work with a class largely populated by students who have not done well in school have lower expectations for their students' achievement (Mallery & Mallery, 1999; Tomlinson, 2006). Teachers who teach these lower track classes place their emphasis on classroom management, memorizing, and drill and practice. This leads to a

situation which Haberman (1991) called a "pedagogy of poverty" where students in honors classes are consistently engaged in higher cognitive tasks and those in remedial classes are less engaged and exposed to instruction characterized by less challenging assignments (Ansalone, 2009; Gamoran, et al., 1995; Haberman, 1991; Hoffer & Gamoran, 1993). Providing an enriched instructional experience to upper track students while denying access to the like by students in the lower tracks leads to the violation of the notion of equality of educational opportunity (Ansalone, 2009; Burris & Wellner, 2005). Tracking produces accumulated incremental effects, which in turn facilitate long-range negative outcomes on the cognitive development of students placed in the lower or regular track (Ansalone, 2006). Students in the upper track classes generally receive more of the intended curriculum while having it presented in a manner that requires students to access higher-levels of thinking (Carbonaro, 2005; Gamoran & Carbonaro, 2003). Rates of student discourse are also higher in the honors courses, which further contribute to the learning gap between the two ability groups (Gamoran et al., 1995). The practice of tracking in our schools represents the continuation of separate but equal where the students' educational experiences in the lower tracks are separate but certainly not equal (Ansalone, 2006; Cooper, 1996: Gamoran & Carbonaro, 2003). Even with recent school reform legislation, schools' efforts have been insufficient to overcome the existing inequalities in our educational system (Wells & Oakes, 1996).

In addition to instructional inequalities, tracks have the potential for creating other disparities. A labeling effect may result which can advance the learning for the high-end student while hindering the achievement for those in the lower tracks (Ansalone, 2009; Oakes & Guiton, 1995). Tracking also positively reinforces the self-concept of high ability students while reducing the self-concept of students assigned to the lower track (Ansalone, 2003, 2006;

DiMartino and Miles, 2005; Gamoran & Berends, 1987; Oakes, 1985). The tracking effects of labeling and on self-esteem demonstrate that tracking does not just contribute to the separation of students along lines of instructional opportunities, race, ethnicity and poverty, but along social and emotional lines as well (Ansalone, 2009; DiMartino & Miles, 2005). The practice of tracking polarizes students into pro- and anti-school camps, creates a system of elite and struggling students, sets lower expectations for the teachers of the lower track, increases time spent on management issues, and encourages stereotyping and segregation (DiMartino & Miles, 2005).

In addition to the abundance of research deploring the practice of tracking, a growing number of studies support the practice of detracking. The notion of detracking is to provide all students—the bottom and the top—access to first class learning opportunities. The theory is if you increase learning opportunities for the low track while continuing to provide quality instructional opportunities to the high track you will decrease the persistent achievement gap while increasing the academic achievement for all students (Burris & Wellner, 2005; Freedman, Delp & Crawford, 2005; Gamoran, 1993).

Many studies indicate that students in lower-track classes tend to learn less than comparable students in higher track classes (Carbonaro & Gamoran, 2002; Gamoran & Mare, 1989). Burris and Welner (2005) did a study in a diverse suburban district in New York. After universally accelerating all students by placing them in detracked classrooms, the percentage of African American and Hispanic students passing the first math Regents exam more than tripled from 23% - 75% (Burris & Welner, 2005). This spike in scores of black and Hispanic students dramatically closed the achievement gap by the time the cohort graduated.

Positive for both high and low achievers

Heterogeneous grouping and differentiated instruction create an atmosphere of equality and caring in the classroom, which then presents students with a better opportunity for success (DiMartino & Miles, 2005; Tomlinson, 2006). Expectations for competence can be treated in such a way as to increase the participation of the lower ability students without depressing the participation of high ability students (Cohen & Lotan, 1995). Having students in heterogeneous ability groups was not a determining factor in students' learning efficacy. Instead, the quality of group processes played the pivotal role and both high and low achievers benefited when the provided group processes were of high quality (Cheng, Lam, & Chan, 2008). In a study of an ethnically diverse, heterogeneously grouped, 8th grade English class Freedman, Delp, & Crawford (2005) found that all students, regardless of probable track placement, made statistically significant gains in writing. Freedman et al., (2005) also found that the student scores increased in a way that resulted in considerably smaller differences in the scores of the lower- and higher-functioning students. A meta-analysis by Lou et al., (1996) analyzed 20 independent findings from 12 studies that directly compared homogeneous grouping with heterogeneous grouping. Student low achievers benefited from placement in mixed-ability (heterogeneous) groupings while high achievers performed equally well in either homogeneous or heterogeneous groupings. In a study in an urban California high school, Cooper (1996) collected student survey data. The researcher found that in the detracked 9th grade English and social studies classes, students reported a positive, intellectually rich, learning environment with equitable learning opportunities. The majority of students surveyed indicated that their detracked classes not only intellectually challenged them, but did so with a culturally sensitive and relevant curriculum (Cooper, 1996).

Positive for low achievers but not for high achievers

Lower-ability students who are grouped in heterogeneous dyads interacted more and completed instructional tasks more efficiently while higher ability students did so more efficiently in homogeneous pairings (Fuchs, Fuchs, Hamlet, & Karns, 1998; Hooper & Hannafin, 1991). Findings such as this lead to a common perspective expressed among teachers that the more able students cover less information, and are not challenged academically when grouped with less able peers. This lack of challenge results in reductions in both efficiency and magnitude of learning (Hooper & Hannafin, 1991). This study and others lead to the perception of many that mixed-ability grouping may support the lower achiever but harm the high achiever.

Hindrances to detracking efforts

Educators implementing detracking efforts in their schools encounter cultural and political obstacles. Often, deeply held beliefs and ideologies about intelligence, racial differences, social stratification, and privilege are confronted. This manifests itself in the important role of institutional culture in our school system. The idea that schools are neutral, nonpolitical places that are in the business of educating students is idealistic. The political and social cultures of the communities in which schools are located make fundamental change very difficult (Cooper, 1996; Oakes, Wells, Jones & Datnow, 1996). Promising efforts toward high standards for all students are frequently cut short by the community's fears that the advantages of high achieving students would be compromised (Oakes & Wells, 1998). Wealthy, White parents want to maintain separate and unequal classes for their children, leaving non-white and poor children in classes that are, by definition, less challenging (Wells & Oakes, 1996). Many supporters of a tracked system are opposed to detracking due to their fear that the behavior of low track students will impede their own student's progress. They cite the lower-track students'

lack of motivation to learn, propensity to act out and their lack of commitment to school (Wells & Oakes, 1996).

Another hindrance is that detracking is not as simple as just changing the school's scheduling or course offerings. Structural or technical changes employed by schools such as changing the master schedule or employing cooperative grouping is not sufficient. To create heterogeneous classrooms in which all students have equal opportunities to succeed and access to high quality instruction attention must be given to the culture and beliefs of the school and community (Cooper, 1999; Yonezawa, et al., 2002). A popular move made by schools that attempt to detrack is the concept of *freedom of choice.* That is, students are able to self-select into the honors grouping or track regardless of prior academic achievement. In a study of four middle schools and six racially mixed high schools, Yonezawa et al., (2002) found that attempts at detracking merely by allowing student choice were unsuccessful. Most of the students in the lower track resisted taking higher track classes because of their prior track placement and their own conceptions of their place in the education hierarchy. Therefore, for detracking efforts to be successful, schools must take into account the omnipresent cultural and social forces as well as pedagogical implications.

Table 2

Sample of examined studies the results of which support detracking

	Agee, 2000	Boaler, 2007	Cooper, 1996	Cooper, 1999	Mosteller et al. 1996	Rubin, 2003
Grade Level/Subject of Study	8th & 10th grade math	High School math	High school English and SS	9th grade English and SS	MS and HS	9th English and history
Sample Size	18 teachers/5 schools	4-year study 3 US HS	319 students	1,090 students	10 experiments (lit. review) each of one school. 2600 students in all	20 students 5 students served as the focal group
Design and Methods	Qual./ Case stud. (Obs./Intvws)	Qual./Quan., (Assessments/Obs. Intvws, Qs.)	Qualitative (Intvws)/Quan. (surveys) Case Study	Quan.(Qs of students)/Qual. (surveys of teachers)	Lit Review	Qual. Ethnographic case study (intvws, notes, obs.)
Findings (Effect Sizes)			Alpha .61-.86	Cronback alpha of .58-.85	-0.33 - + 0.29	
Strengths	3 methods of data collection	4-year longitudinal study		Large student sample size	Large sample	
Results Trusted or Tentative	Tentative—a glimpse of a few teachers each in their own context.	Trusted-Conclusions were realistic considering limitations	Tentative. Randomized studies are preferred to matched pairs.	Trusted	Trusted—well designed methodology. Sig. ES	Tentative—low sample size
Findings	It is reasonable to ask teachers to develop literature instruction that provides students of differing abilities to have success. Stugglers can have success when appropriate texts are used and ways of reading are encouraged.	Schools which achieved a high level of learning in mixed-ability classrooms attended to the complexities of racial relations. Curriculum is only one part of a complex interconnected system.	The core program (mixed ability groups) reported a positive learning environment with equitable learning opportunities that was intellectually rich. Although results were positive this study highlighted that difficulties with detracking are sometimes too great an obstacle to overcome.	Heterogeneous grouping (core) promoted greater academic achievement for more students while maintaining standards for advanced students.	Five studies favored skill grouping, three favor whole-class (hetero) grouping, two had effect sizes near zero.	Structural detracking alone is not enough. Must bridge the gap between teacher practices and students social worlds. Detracking reform holds promise—communicates to students that school is committed to equality.

International Research on Tracking

Tracking or streaming has been a common practice in Great Britain for over half a century. It was first introduced in reaction to the 1931 *Primary* School *Report* as a means for classroom management and to respond to a greater concern for excellence rather than equity (Ansalone, 2003). Global educational policy is now focused on raising standards. In England and Wales, this movement has encouraged schools to consider how best to organize pupils and teachers to raise achievement on national tests and examinations (Ireson & Hallam, 2005). As in the United States, the establishment of a national curriculum and the accompanying desire for academic excellence have led to a resurgence in streaming and setting. Most secondary schools in the UK employ some form of ability grouping—usually setting—for at least a portion of the school day (Hallam & Ireson, 2007). The parents of high achievers particularly are in favor as they see these types of ability grouping as an academic benefit for their children (Ansalone, 2003). As in the United States, British researchers have found that low-ability streams includes a disproportionate number of pupils who are ethnic minorities and are of low socioeconomic status.

Ireson and Hallam (2005) studied primary and secondary schools throughout the UK. They examined the effects of ability grouping on a wide range of student outcomes including a student's liking for school, and his or her perceptions of teaching. Using a questionnaire, they found that grade 7-9 students in schools with setting liked school less than pupils in mixed-ability grouped schools did. Furthermore, students in low ability groups liked school less than students in the higher ability groupings did ($F_{2;\,1427} = 9{:}24$, $p < 0001$, ES ¼ 0:01). In terms of students' perceptions of teaching, the students in the higher sets were significantly more positive

($F_{2;\,2184}$ = 14:0, p < 0001; ES ¼ 0:01). In summary, attention to the affective consequences of ability grouping should be studied further and considered.

Ireson, Hallam, and Hurley (2005) conducted a quantitative study to determine the effects of setting on student achievement on the General Certificate of Secondary Education (GCSE) examinations taken in grade 11. They used multi-level modeling and controlled for Attendance, SES (i.e., socio-economic status), Gender, and prior attainment. Their research involved over 6000 students in 45 comprehensive secondary schools (Ireson et al., 2005). They found that prior attainment in the subject was a strong contributor to their GCSE exam scores. The number of years students were placed in different ability groups had no effect on their scores on the GCSE. These findings are consistent with Slavin (1990) who found that if students are subject to the same curriculum, ability grouping has little effect on student achievement. It is noted that curriculum differentiation in English schools is not as pronounced as in American schools (Ireson et al., 2005). This likely explains the disparity in student achievement between the different groupings in American schools.

Kerckhoff (1986) discussed the two theories behind ability grouping; the traditional and divergent theories. The traditional theory suggests that ability grouping is good for all students while the divergent theory espouses that ability grouping accelerates the achievement of the high end while depressing the achievement of the low. Comparing mean exam scores and using a multiple regression model Kerckhoff (1986) found the divergent theory was supported. Students in the lower ability (remedial) group lost the most ground while those in high ability groups increased their performance level beyond comparable students in ungrouped school settings. Although the British have two types of ability grouping—streaming and setting—Kerckhoff was not able to study them independently. This is an experimental design flaw, which the researcher

noted, as several researchers have found that these two grouping practices have pronounced effect differences (Hallinan & Sorensen, 1983; Oakes, 1992).

Wiliam and Bartholomew (2004) did a study on the effects of setting on math achievement in British secondary schools. The researchers followed a cohort of 709 secondary school students in the greater London area as they moved from year 8 to year 11. They collected data from national exams at the end of years 9 and 11. The article reported the effects of setting on the cohort's progress as measured by the two exam scores. The relationship between the grade 9 (Key Stage 3) exam score and the grade 11 (GCSE) score for the four sets (top, upper, lower, bottom) in each school was determined. Students in top sets achieve over half a grade (i.e. 0.58 grades) higher on the GCSE than would have been predicted from their Key Stage 3 scores, while those in the bottom sets scored just over half a grade (0.51 grades) lower than predicted from their Key Stage 3 scores. This study on student math placement replicates findings from other studies that grouping students has the effect of advancing the learning of those in the high group while slowing the learning pace of those in the low group (Gamoran & Mare, 1989; Gamoran et al., 1995; Hoffer, 1992; Kerckhoff, 1986). Interestingly, the smallest differences between schools on the GCSE were at schools where teachers continued to utilize small groups and individualized work as pedagogical strategies. Therefore, the effects of setting may not be due to the grouping of students by ability but instead by the teacher's instructional practice. Wiliam and Bartholomew (2004) noted, teachers who taught mixed-ability classes used a greater variety of instructional strategies and were "better teachers" even though they did not enjoy teaching such classes (p. 290).

Schofield (2010) did a meta-analysis of research on tracking in other developed countries. She considered two fundamental questions: 1) "Is having higher achieving

schoolmates/classmates commonly associated with larger achievement gains for secondary school students?" and 2) "Is ability grouping with curriculum differentiation commonly associated with a larger achievement gap for secondary school students?" (p. 1492). She also looked at three possible mediators of ability grouping's effects: "increased social class homogeneity, more effective pedagogical behaviors in classes with higher achieving students, and more positive peer influences in higher achieving schools and classrooms" (Schofield, 2010; p. 1494). The question of whether the mean ability level of classmates affects student academic achievement gains is important in understanding the effects of ability grouping. High quality studies in secondary schools are necessary to establish this relationship between mean peer ability levels and student achievement. Schofield found research from schools in the U.S., Belgium, Canada, France, and New Zealand that shows increasing the mean of a class's initial achievement significantly increases achievement gains for most students. The results "quite consistently suggest that low-achieving students are more positively influenced by high-achieving classmates than are high-achieving students" (p. 1500). These studies conclude that overall academic achievement is greater when the range of abilities within the classroom is large. The lingering question remains: Does this grouping of students with a wide range of abilities hurt the high-end student?

In a study of student achievement before and after tracking, Venkatakrishnan and Wiliam (2003) found that a greater number of students gain in mixed-ability classes compared to gains when high achievers are placed in a separate track. They also found that the high achievers gained somewhat more in homogeneous classes than in mixed-ability ones. This is consistent with Kulik and Kulik (1992) who found that gifted students who were tracked outperformed students of similar ability who were not tracked. Again, this is the crux of the present

dissertation. Students in high-ability groups are provided with an enriched curriculum and a greater opportunity to learn. Consequently, "students in the lower tracks suffer from the loss of intellectual stimulation generally associated with students possessing more social capital" (Ansalone, 2003; p. 8). However, if both groups have the same enriched curriculum with high quality differentiated instruction, do high achieving students benefit from being separated from the lower achievers?

Synthesis and significance of reviewed research

Studies can be segregated into those that show: positive effects of tracking on the high achieving group, positive effects of detracking on the low achieving group, positive effects of detracking on the high achieving group, negative effects of tracking on the low achieving student group, and negative effects of detracking on the high achieving student group. Although there are studies that fall in categories other than these, the overwhelming number of studies found for this researcher's literature review fall in one of the five aforementioned categories (see Table 3).

The practice of tracking in America's schools continues today despite evidence that the practice does not benefit students in the lower tracks or courses. Students who are placed in the lower courses tend to remain at this level or track throughout their years of compulsory education. Students who populate these lower tracks disproportionality come from low socio-economic families, and are either black or Hispanic. The effect on students who become stuck in these lower tracks is that they fall further and further behind their higher tracked peers on measures of student achievement. This separation between poor, black, and Hispanic students and their white and Asian peers contributes significantly to the gap in their quantified academic achievement levels. Tracking remains a practice today, largely because the practice has been found to benefit those students grouped with their higher performing peers. Students in higher

tracks do better than students in lower tracks. This illustrates the dilemma that has faced educators for over a century. How do we serve the academic needs of all students? To do so requires balancing the desire for a challenging and rigorous education for the high achiever while providing a similarly rigorous and valuable experience for the lower achiever.

Table 3

Synthesis of reviewed literature

	Positive Effects of Tracking High Achiever	Positive Effects of Detracking Low Achiever	Positive Effects of Detracking High Achiever	Negative Effects of Tracking Low Achiever	Negative Effects of Detracking High Achiever
Qualitative Studies	Ansalone, (2010) Carbonaro, (2005) Gallagher, (1997) Gamoran, (1986) Gamoran, (1989) Gamoran & Mare, (1989) Hoffer & Gamoran, (1993) Kulik & Kulik, (1982) Perrone et al., (2010) Rogers, (1993) Rogers, (2007) Shields, (2002) Wells & Oakes, (1996)	Anaslone, (2006) Argys, Rees & Brewer, (1996) Boaler, (2007) Burris & Wellner, (2005) DiMartino & Miles, (2005) Oakes & Wells, (1998)	Freedman et al., (2005)	Ansalone, (2009) Ballon, (2008) Carbonaro, (2005) Chambers, (2009) Cooper, (1996) DiMartino & Miles, (2005) Gamoran, (1989) Gamoran & Mare, (1989) Hoffer & Gamoran, (1993) Hooper & Hannafin, (1991) Tomlinson, (2006) Wells & Oakes, (1996)	Argys, Rees & Brewer, (1996) Benbow & Stanley (1996) Gamoran, (1989) Oakes & Wells, (1998) Preckel, Gotz & Frenzel, (2010)
Qual./Quan.	Gamoran et al., (1995) Gamoran, (1993) Gamoran & Carbanaro, (2003)	Cheng, Lam, & Chan, (2008)	Cheng, Lam, & Chan, (2008)	Gamoran et al., (1995) Gamoran, (1993) Gamoran & Carbanaro, (2003)	Benbow & Stanley, (1996)
Experimental w/Control Quasi-Experimental	Hooper & Hannafin, (1991) Goldring, (1990) Allan, (1991)	Hooper & Hannafin, (1991)			Hooper & Hannafin, (1991)
Ave. Effect Size	Greater than .25	Greater than .25	Greater than .25	Greater than .25	Greater than .25

Gap in research

The benefits of homogeneous grouping for the higher achiever are usually quantified in the research by comparing students between high tracks and low tracks. Slavin (1995) found

these studies problematic. Slavin suggested studies that comparisons of tracked students to non-tracked students are far more meaningful. This statement begins to reveal a gap in the current research on tracking/detracking. In her doctoral dissertation, Livingston (2010) stated "there appears to be a gap in the research that explores the impact of tracking on students with similar test scores but who are placed in different tracks" (p. 3). Schofield (2010) identified a gap in the research when discussing that ability grouping and curriculum differentiation are commonly intertwined. Therefore, most research is an examination of their combined effect rather than their effects independently. A common methodology in U.S. tracking studies is to compare the progress of students in advanced classes with those in lower tracks. According to Sternberg (1985), this comparison is problematic because student achievement gains might have been identical in heterogeneous classrooms with a common curriculum. Schofield (2010) acknowledged that one approach used to minimize methodological problems is to compare the progress of students in schools that have ability grouping with curriculum differentiation with schools that do not.

Schofield (2010) also found that many studies of ability grouping and individual achievement gains do not use measures of curriculum differentiation as a control variable "experimental studies of the impact of ability grouping alone that control for the effects of curriculum differentiation are rare beyond ninth grade" (p. 1499). Schofield (2010) found quite a few studies on the combined effects of ability grouping and curriculum differentiation on achievement but few that determine the effects of ability grouping alone. This dissertation intends to fill this gap.

Quantitative studies on the effects of detracking are inconclusive. Researchers claim both positive (Agee, 2000; Alvarez & Mehan, 2005; Slavin, 1991, 1995) and negative effects

(Allan, 1991; Benbow & Stanley, 1996; Kulik & Kulik, 1982). Furthermore, much of the qualitative research on detracking focuses on school, community, and affective factors rather than its effects on student achievement (Oakes et al., 1997; Rubin & Noguera, 2010). Tieso (2003) stated "there have been few controlled studies of the effects of different types of grouping arrangements on student achievement" (p. 32). Many of the researchers have focused on analyzing the dynamics at play in our schools and communities. As a result, these studies challenge deeply held beliefs and ideologies that are at the heart of our educational system (Cooper, 1996; Rubin, 2003). Gamoran, (2009) called for new research that may capture the benefits of detracking and differentiation. He stated "ultimately, how students are arranged matters less than the instruction they encounter, so bringing together research on tracking with research on teaching offers the most useful way to continue to shed light on this topic of continuing interest" (p. 15).

One group likely to oppose the dismantling of the tracking system, are the parents of students who were previously placed in the higher track classes (Ansalone & Biafora, 2010; Oakes, et al., 1997; Rubin, 2006). These parents fear that efforts to promote detracking will result in lower academic standards for their students (Ansalone, 2003). This study will be an examination of whether this fear is justified. Does the placement of honors English students into heterogeneous (i.e., mixed ability) classrooms affect their performance on measures of academic achievement when compared to their homogeneously grouped peers?

CHAPTER THREE

Methodology

Methodological orientation. The number of research studies and articles on the practice of ability grouping (i.e., like-ability) and/or (i.e., mixed-ability grouping) is voluminous. Some of the researchers note the benefits of tracking while others cite significant issues as the result of the practice. Many of the studies are qualitative explorations of the practice of tracking or ability grouping by comparing the perceptions or affective components of stakeholders (Ansalone & Biafora, 2010; Argys et al., 1996; Biafora & Ansalone, 2008; Gallagher, 1997; LaPrade, 2011: Preckel et al., 2010). The quantitative studies are fewer and tend to be comparisons of progress of students in the low tracks (groups) with those in the high tracks (Gamoran, 1989, 1993; Hoffer & Gamoran, 1993). A good number of studies are in relative agreement. Tracking is good for the high end student (Fuchs et al., 1998; Goldring, 2001; Hattie, 2002; Kulik & Kulik, 1982; Preckel, et al., 2010; Rogers, 2007; Shields, 2002; Van Tassel-Baska, 2005) but detrimental for the student at the low end (Ansalone, 2000, 2004, 2009; Archbald & Keleher, 2008; Chambers et al., 2009; Gamoran et al., 1995; Mallery & Mallery, 1999; Mickelson, 2001; Oakes et al., 1992). Add to these findings that students of color and poor students make up a disproportionate number of students in the low tracks and our educational system and our country are in a moral quandary (Burris & Wellner, 2005; Chambers, 2009; Gamoran & Mare, 1989; LaPrade, 2011; Slavin 1995). How does our educational system best provide a high quality, rigorous, and valuable education to its students who demonstrate a wide range of academic skill and ability levels?

Research design

This study follows a non-experimental (i.e., *ex post facto*), correlational, evaluative, explanatory associational design. Gay et al, (2009) claimed that a study is correlational when data are collected "to determine whether, and to what degree, a relationship exists between two or more quantifiable variables" (p. 196). The objective of this study is to determine to what degree ability grouping high school honors English students is correlated with student reading and/or writing achievement.

Being an *ex post facto* research study, the manipulation of the independent variable or alleged cause (i.e., ability grouping) and the studied effect (i.e., student achievement) was not possible. The groups of students in this study were not randomly assigned so establishing the groups as initially similar or determining if a statistical manipulation is needed is an important exercise (Wright, 2006). The ideal research design would require random assignment of students but such designs are impractical because it is not ethical to force students into different educational groups for experimental convenience (Swiatek & Benbow, 1991). These ethical concerns make quasi-experimental or non-experimental designs better (Cook & Campbell, 1979; Swiatek & Benbow, 1991). This study also is considered a form of applied research for it is evaluating the theory that grouping students homogeneously according to ability has a positive influence on student academic performance. It is evaluative because it involves collecting and analyzing data about the effectiveness of a particular practice, in this case ability grouping. This study is intended to answer—among others—the question: Does homogeneously grouping honors English students by ability have merit in terms of better preparing them for state and pre-college assessments of reading and writing? Therefore, using these research designs, the researcher was able to test questions concerning the influence of heterogeneous or homogeneous

student ability grouping (i.e., independent variable) on students' performance on a state reading assessment and scores on the reading and writing portion of the PSAT (i.e., dependent variables). Gay et al. (2009) maintained that since the independent variables are not manipulated by the researcher they are better referred to as the *grouping variable*. This researcher acknowledges the point and will used the two terms interchangeably. See Figure 1 and 2 for a comprehensive view of the research design.

School # 0	Year	School # 1	Year
Student Baseline Data 8th grade WASL Scores in Reading ↓ Grouping Variable 9th grade & 10th grade—Honors English w/ Ability Grouping (Homogeneous) ↓ Dependent Variable (Student Achievement) HSPE Reading: Spring of Soph. Yr. PSAT (R & W): Fall of Soph. Yr. PSAT (R & W): Fall of Junior Yr.	Spring 2009 2009- 2010 2010- 2011 2011 2010 2011	Student Baseline Data 8th grade WASL Scores in Reading ↓ Grouping Variable 9th grade & 10th grade—Honors English w/ Ability Grouping (Heterogeneous) ↓ Dependent Variable (Student Achievement HSPE Reading: Spring of Soph. Yr. PSAT (R & W): Fall of Soph. Yr. PSAT (R & W): Fall of Junior Yr.	Spring 2009 2009- 2010 2010- 2011 2011 2010 2011
Dependent Variable: Student Achievement—HSPE, PSAT. Independent Grouping Variable: Type of Ability Grouping - Homogeneous Controls: Student: Gender, Ethnicity, SES, ELL, Special Education, Attendance, grade 8 assessment results		Dependent Variable: Student Achievement—HSPE, PSAT. Independent Grouping Variable: Type of Ability Grouping - Heterogeneous Controls: Student: Gender, Ethnicity, SES, ELL, Special Education, Attendance, grade 8 assessment results	

Figure 1. Review of research design: Cohort 0

School # 0	Year	School # 1	Year
Student Baseline Data 8th Grade MSP Scores in Reading ↓ Grouping Variable 9th Grade & 10th Grade—Honors English w/ Ability Grouping (Homogeneous) ↓ Dependent Variable (Student Achievement): HSPE Reading: Spring of Soph. Yr. PSAT (R & W): Fall of Soph. Yr. PSAT (R & W): Fall of Junior Yr.	Spring 2010 2010-2011 2011-2012 2012 2011 2012	Student Baseline Data 8th Grade MSP Scores in Reading ↓ Grouping Variable 9th Grade & 10th Grade—Honors English w/ Ability Grouping (Heterogeneous) ↓ Dependent Variable (Student Achievement): HSPE Reading: Spring of Soph. Yr. PSAT (R & W): Fall of Soph. Yr. PSAT (R & W): Fall of Junior Yr.	Spring 2010 2010-2011 2011-2012 2012 2011 2012
Dependent Variable: Student Achievement—HSPE, PSAT Independent/Grouping Variable: Type of Ability Grouping - Homogeneous Controls: Student: Gender, Ethnicity, SES, ELL, Sped., Attendance, Grade 8 assessments		Dependent Variable: Student Achievement—HSPE, PSAT Independent Grouping Variable: Type of Ability Grouping - Heterogeneous Controls: Student: Gender, Ethnicity, SES, ELL, Sped., Attendance, Grade 8 assessments	

Figure 2. Review of research design: Cohort 1

Data Collection

Two cohorts of students are the subject of this study. Each cohort is, in effect, its own study. Students from each school who completed the 8th grade Washington Assessment of Student Learning (WASL) in reading and subsequently completed honors English in grades 9 and 10 in each of their schools are the focus of this research. Both high schools in this study allow incoming grade 9 students to self-select the honors English curriculum. The WASL in reading was given in grade 8 (Spring 2009) and the MSP was given in grade 8 (Spring 2010). Both will be used as a baseline for their respective cohorts. The WASL and MSP report scores

on a scale of 1 to 4 (1 = below basic, 2 = basic, 3 = proficient or meets standard, 4 = advanced or exceeds standard). There are also raw scores given to each student where 1 (below basic) = 250-374, 2 (basic) = 375-399, 3 (meets standard) = 400-418 and 4 (exceeds standard) ≥ 419. Student scores on the grade 10 High School Proficiency Exam in reading, which is taken in the spring of their sophomore year and is reported on the same 4-point scale, is a dependent variable. PSAT scores of the students also will be collected (Cohort 1: fall 2010 and 2011, cohort 2: fall 2011 and 2012). To comply with the contention of Gay et al. (2009), that every effort be made to compare groups that are as equivalent as possible, a chi square analysis on the independent variables of Gender, SES, ELL, Ethnicity, and Special education will be completed.

According to the *Washington State Assessment Coordinator's Manual* (2012) a concerted effort is made to ensure "the assessments show respect for cultural diversity and are not biased in a way that would affect the performance of particular groups of students" (p. 2). To alleviate reliability and validity concerns, all students in grades 3–8 and 10 are required to participate in all grade-level testing. Due diligence was practiced by all district staff to follow the test administration protocols as outlined in the *Washington State Assessment Coordinator's Manual* for each year of scheduled state testing. All Washington State score reports were coded to maintain confidentiality. The results pertaining to the dependent variables were then organized into cohort groups as follows:

- Cohort #1: 2009 - 8th grade WASL / 2011 10th grade HSPE / 2010, 2011 PSAT.
- Cohort #2: 2010 - 8th grade MSP / 2012 10th grade HSPE / 2011, 2012 PSAT.

Population

Two suburban public high schools in Washington State, in the same large, very diverse school district, with similar demographics, will be involved in the study. The district has

approximately 27,000 students in 28 elementary schools, 6 middle schools, 4 comprehensive high schools, and 3 alternative/choice schools. The district's demographics have the approximate values of 44% Caucasian, 12% African American, 17% Asian, 17 % Hispanic, 18% Asian /Pacific Islander, and 7% multi-racial. The free and/or reduced lunch percentage of the district is approximately 50%. The two high schools in this study each have similar demographics, which approximate the district's demographics. Each of the chosen high schools has students attending grades 9-12 and provides their students with a district adopted English curriculum. School 0 provides students the honors curriculum in the traditional model where students are grouped homogeneously according to ability. School 0 has honors English students in classrooms separate from their non-honors English peers. School 1 provides the honors curriculum to students who are grouped heterogeneously according to ability. Therefore, in School 1, approximately half of the students in a particular classroom have taken honors English prior to the ninth grade while the other half have not. Therefore, in School 1, honors English students are in the same classroom as students who previously completed a lower level English curriculum. In these classrooms, teachers differentiate their instruction to meet the academic needs of each type of student—core or honors. Students who are provided and complete honors-level work are awarded honors English credit on their high school transcript. Students will be identified as honors or core (i.e., non-honors) by this transcript designation.

Sample Size

Cohort 0 includes 388 students who took honors English in grades 9 and 10 during the 2009–2011 school years. From School 0, 235 students and from School 1 140 students took the 8^{th} grade state reading assessment (WASL). From School 0, 240 students and from School 1 144 students completed the 10^{th} grade state reading assessment (HSPE). From School 0, 225 honors

English students and from School 1, 96 students took the 2011 PSAT in the fall of their junior year. Only 2 students completed the reading and writing portions of the 2010 PSAT from School 0, with 36 completing it from School 1.

Cohort 1 includes 474 students who took honors English in grades 9 and 10 during the 2010–2012 school years. From School 0, 276 students and from School 1, 165 students took the 8th grade state reading assessment (MSP). From School 0, 294 students and from School 1, 175 students completed the 10th grade state reading assessment (HSPE). From School 0, 285 honors English students completed the 2011 PSAT Reading and writing sections while 148 completed the 2012 PSAT Reading and writing sections. School 1 had 168 honors English students complete the 2011 PSAT Reading and writing sections while 87 completed the 2012 PSAT Reading and writing sections.

Procedures

A written proposal was presented to the district director of student assessment prior to the commencement of this study. Following an in depth discussion, permission was granted by the school district superintendent. Participant groups subsequently were identified utilizing *Skyward*, the district student management software package. Participants who met the following criteria were chosen to participate in this study: (a) each student will be in the same school for both their 9th and 10th grade year and (b) each student in the sample will have completed honors English in both their grade 9 and grade 10 years.

Two cohorts of students from School 0 and School 1 will be involved in the study. The first cohort is composed of students who were continuously enrolled in honors English 9 during the 2009–2010 school year and honors English 10 during the 2010–2011 school year. The second cohort is composed of students who were enrolled continuously in honors English 9

during the 2010–2011 school year and honors English 10 during the 2011–2012 school year. School effect for each cohort was taken into account by performing a chi square analysis for both School 0 and School 1. The chi square analysis was conducted controlling for the variables of English Language Learner (ELL), Special education, free-reduced lunch (FRL), Gender, Ethnicity, and full-day Absences for each honors English student enrolled in School 0 and School 1. In addition, a *t*-test was performed on the independent variables of full-day absences and grade 8 state reading scores in order to establish the similarities of School 0 and School 1.

Data analysis

A multiple regression analysis with the predictors: Grade 8 WASL (Cohort 0) or MSP (Cohort 1), student Gender, FRL, Ethnicity, Special education, student Absences, ELL, and School was utilized. The significance and magnitude of the predictor variable, School, which is a proxy for "homogeneous ability grouping" or "heterogeneous ability grouping" will then be analyzed.

Without random assignment, the groups being studied are more likely to differ on some important variable other than the variable under study (Gay et al., 2009). A way to control extraneous variables is to compare groups that are similar (Gay et al., 2009; Spector, 1981). A chi-square analysis and *t*-tests will be run to determine the likeness between the two student samples from each school. Even though the researcher of this study was unable to manipulate class grouping, the aforementioned controls coupled with the chi-square analysis and t-tests act to measure student similarity between the schools of both cohorts.

A non-experimental (i.e., *ex post facto*), explanatory associational design was utilized to gather and analyze data in this study. An analysis of the data was conducted to determine the influence of student grouping (homogeneous vs. heterogeneous) in high school English classes

on honors students' academic achievement in reading and writing. Washington State assessment scores in reading at grades 8 and 10 and PSAT Reading and writing scores at grade 10 and 11 were analyzed using SPSS version 20. Pre-treatment scores were measured utilizing the 8th grade reading WASL (Cohort 0) or MSP (Cohort 1), which was administered to each cohort in the spring of their 8th grade year.

Gay et al. (2009) posit that when random assignment is not possible there are many quasi-experimental designs from which the researcher may choose. Educational research rarely lends itself to large-scale experimental design and true randomization (Grunwald & Mayhew, 2008). In fact, Luellen, Shadish, and Clark (2005) believed these designs reflected better the complexity inherent in the educational context. To account for threats to validity, the quasi-experimental researcher "should make every effort to include groups that are as equivalent as possible" (p. 259). Because groups were not aware that they were involved in a study, possible effects from reactive arrangements are reduced. This research was casual-comparative in nature since it involved two groups of participants (like-ability and mixed-ability) and one dependent variable (student achievement). The design is *ex post facto* since the effects of the independent variables on the dependent variable are studied in retrospect. In such a design, it is critical to match the subjects on several critical variables. Since this study involves two or more measurement periods, the chi-square and *t*-tests can be used to test for differences among treatment groups (Spector, 1981). The study is also correlational. As stated by Gay et al. (2009), correlational research "involves collecting data to determine whether and to what degree a relationship exists between two or more quantifiable variables" (p. 196). Furthermore, this study is a type of evaluation research because it is an effort to answer whether a type of practice

(homogeneous ability grouping) is beneficial to the high achiever for high school English instruction (Gay et. al., 2009).

Multiple regression analysis is appropriate for associational research because it allows the researcher to determine how much of the variance in the dependent variable is attributable to the independent variables (Gay et al., 2009). The research design used in this study met the conditions for multiple regression. The dependent variable is a normally distributed scale variable and independent variables are scale or dummy coded variables (Leech, Barrett & Morgan, 2011). These statistical tests provide β, F, and R^2 values as well as significance levels (p), all important statistical measures in explaining relationships between variables. The results of all tests will be discussed in Chapter IV. The statistical measures will be used to determine student demographic similarity prior to students entering homogeneous or heterogeneous classrooms for honors English instruction. The analyses will then be used to determine the degree to which the independent variables influenced student academic achievement as measured by the dependent variables.

The demographic independent variables will be entered into the regression to determine their specific influence on the dependent variables of student scores on state and pre-college assessments of reading and writing achievement. It is acknowledged by social science researchers that demographic variables such as Gender, SES (measured in this study by percent free-reduced lunch), and Ethnicity have a significant effect on student achievement (Caldas, 1993; Sutton & Soderstrom, 1999). The independent variable of student attendance also has been found to be moderately to strongly correlated with student achievement measures (Sutton & Soderstrom, 1999). Therefore, these variables will be included in the regression models to allow for their statistical removal. The goal of this study is to determine whether ability grouping has

an effect on high school honors English students' scores on measures of reading and writing achievement.

Research question

The significance level of this study was at the .05 probability level or higher. The present study is enveloped in the aforementioned conceptual framework and is guided by one overarching and four subsidiary research questions. Specific and intentional outputs of SPSS version 20 will be used to answer the following research questions:

- To what extent, if any, does ability grouping of high achieving students defined as 9th and 10th grade honors English students in a suburban Washington State school district, affect their performance on state and pre-college assessments of reading and writing achievement when controlling for student mutable variables?

Subsidiary research questions. (a) To what extent, if any, does placement of 2009–2011 honors English students in homogeneous ability grouped English classes have on their subsequent performance on state and pre-college assessments of reading and writing achievement? (b) to what extent, if any, does placement of 2009–2011 honors English students in heterogeneous ability groups have on their subsequent performance on state and pre-college assessments of reading and writing achievement? (c) to what extent, if any, does placement of 2010–2012 honors English students in homogeneous ability grouped English classes have on their subsequent performance on state and pre-college assessments of reading and writing achievement? (d) to what extent, if any, does placement of 2010–2012 honors English students in heterogeneous ability groups have on their subsequent performance on state and pre-college assessments of reading and writing achievement?

Variables

This study was designed to determine the extent to which grouping students according to ability and a previously taught honors English curriculum contributes to their performance on state and pre-college assessments of reading and writing achievement. The influence of the predictor variable (homogeneous or heterogeneous ability grouping) on the dependent variable of academic achievement as measured by the grade 10 state reading assessment and the PSAT was examined. The independent variables of student Gender, Ethnicity, SES (free-reduced lunch), Special education, ELL status, and student Attendance will be controlled in the multiple regression.

Instrumentation

In 1993, Washington State embarked on a comprehensive school change effort with the primary goal to improve teaching and learning. To support this school change effort the state legislature created the Commission on Student Learning. This body was charged with three important tasks:

- Establish Essential Academic Learning Requirements (EALRs) which identify what students should know and be able to do.
- Develop an assessment system to measure student progress towards the attainment of these standards.
- Recommend an accountability system that recognizes successful schools and offers supports and assistance to the schools that proved less successful.

The Essential Academic Learning Requirements (EALRs) in reading, writing, communications, and mathematics were adopted in 1995 and revised in 1997. The grade 10 assessments became mandatory in spring of 2000 and the grade 8 assessments became mandatory in spring 2006.

Currently, the Washington State testing program is known as the Washington Comprehensive Assessment Program (WCAP). The state assessments were developed specifically to measure student progress toward meeting the EALRs, grade level expectations (GLEs), and Performance Expectations (PEs). According to the *Washington State Assessment Coordinator's Manual* (2012) the WCAP is a standards-based student assessment program designed to:

- assist schools, districts, and the state in improving student learning;
- report students' level of proficiency relative to the Essential Academic Learning requirements (EALRs) in reading, writing, mathematics, and science;
- measure progress toward district and school improvement targets;
- serve as Washington's accountability measure to meet federal requirements under the *No Child Left Behind Act*; and
- be used as one of the state's requirements for a standard high school diploma, beginning with the graduating class of 2008 (p. 1).

The state assessments require students to select and construct responses in order to demonstrate their mastery of each of the EALRs. The reading test includes multiple-choice, short-answer, and extended response items. The operational test forms are standardized and on-demand; meaning students take the exam at the same time during the school year, under like conditions, and respond to the same items. The reading assessment is untimed and guidelines have been established and accommodations available to facilitate the inclusion of special education students.

Classroom teachers and curriculum specialists throughout Washington State were instrumental in developing the items for the criterion-referenced state assessment. Together with

the Pearson Educational Measurement, these teachers and specialists provided final review and recommendations to approve items after pilot testing. A bias and fairness committee also was established to conduct a sensitivity review of all items. This review was an effort to look for words or content that might have proven offensive or disadvantageous to students unrelated to the skill being assessed. A pool of pilot-tested items was established for each grade level and content area. Selections from the item pool are used to form new forms of the assessment each year. Procedures are applied to statistically equate each test in order to maintain the same year-year performance level standards.

Scoring rubrics, developed by the content committees, were used to score student responses. Statistical analyses based on classical test theory and modern item response theory were done to evaluate item effectiveness and to empirically examine the presence of differential item functioning or item bias. The Washington State assessments (the Washington Assessment of Student Learning—WASL, the Measurement of Student Progress—MSP; and the High School Proficiency Exam—HSPE) report results on a 4-point scale. Level 4 (advanced), Level 3 (proficient), Level 2 (basic) and Level 1 (below basic). All state assessments are scaled so that a scaled score of 400 is the benchmark for being proficient or Level 3. Students must be either Level 3 or Level 4 to be judged as meeting standard. Below, and in Table 4, a breakdown of the descriptions for each level can be found:

Level 4: *Advanced*—Student performance is judged superior, which is notably above that required for meeting the standard:

- grade 8: scale score of 419-500.
- grade 10: scale score of 427-525.

Level 3: *Proficient*—This level represents solid academic performance. Students reaching this level have demonstrated proficiency over challenging content, including subject-matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate for the content and grade level.

- grade 8: scale score of 400-418.
- grade 10: scale score of 400-426.

Level 2: *Basic*—This level denotes partial accomplishment of the knowledge and skills that are fundamental for meeting the standard.

Level 1: *Below Basic*—This level denotes little or no demonstration of the prerequisite knowledge and skills that are fundamental for meeting the standard.

Table 4

Washington State Assessment (WASL, MSP, and HSPE): Scale Score Ranges Reading

	Level 1	Level 2	Level 3	Level 4
Grade 8	250-374	375-399	400-418	419-500
Grade 10	225-374	375-399	400-426	427-525

Washington Assessment of student Learning (WASL) and measurement of student progress (MSP) reading grade 8.

The WASL and the MSP are both criterion-referenced tests designed to measure the extent to which the student has mastered the knowledge and skills identified by state standards. The most current version of the WASL was used to assess students' reading proficiency in grade 8 from the spring of 2006 to 2009. The WASL for grades 3-8 was replaced by the MSP in the spring of 2010.

High school proficiency exam (HSPE) reading grade 10

Like the WASL and MSP, the grade 10 HSPE is a criterion-referenced test developed to measure the extent to which the student has mastered the knowledge and skills identified by state standards. The grade 10 HSPE replaced the grade 10 WASL in the spring of 2009.

Reliability

There are many definitions of reliability, all of which have their origins in Classical Test Theory (Feldt & Brennan, 1989; Webb, Shavelson, & Haertel, 2007). The Classical Test Theory approach builds on the concept of an ideal, error-free, or true measurement score. In Classical Test Theory, it is assumed that each observed score (X) contains a true component (T) and an error component (E). When measuring a construct, unsystematic errors occur (Feldt & Brennan, 1989).

The reliability components examined in the State of Washington's reliability determination were internal consistency, standard error of measurements, rater agreement, and decision consistency and accuracy. Internal consistency reliability is a measure of whether or not students perform consistently across items. Internal consistency can be estimated by Cronbach's coefficient alpha. Cronbach's alpha estimates internal consistency by determining how all items on a test relate to all other test items and to the total test (Gay et al., 2009). Washington State for both the WASL (grade 8) and the HSPE (grade 10) used Cronbach's alpha. There are two requirements to estimate score reliability:

- The number of items should be sufficient to obtain stable estimates of students' achievement.
- All test items should be homogeneous (i.e., similar in format and measure very similar knowledge and skills).

Coefficient Alpha scores above .90 are considered *highly reliable* or *good*, between .80 and .89 are considered to have *good* reliability, between .70 and .79 are considered to have *fair* or *acceptable* reliability, between .60 and .69 are considered to have *marginal* or *questionable* reliability, and coefficients under .60 are considered *poor* or *unacceptable* reliability (George & Mallery, 2003).

Test score reliability is a measure of the degree to which scores on a test truly measure the knowledge and skills of the examinee in relation to the tested knowledge and skills. Using Classical Test Theory, reliability compares observed score variance with true score variance. Several methods are available to estimate score reliability: test-retest, alternate forms, internal consistency, and generalizability analysis are considered the most common (Spector, 1981). Both test-retest and alternative forms reliability estimates require the test taker to engage with the tests for a significant amount of time. This has a potential effect on the examinee in terms of fatigue and loss of motivation (Gay et al., 2009). Due to this, Washington State used internal consistency measures to estimate score reliability for the reading assessment. The WASL and HSPE combine multiple choice, short answer, and completion; therefore, the examinee's performance may differ decidedly from item to item. This heterogeneity of items in the reading assessments may tend to underestimate the reliability of scores as estimated by Cronbach's coefficient alpha. Even with this item heterogeneity, the WASL has an alpha coefficient for reading in grade 8 (2009) of 0.88 and the MSP for reading grade 8 (2010) of 0.88. The Coefficient alpha for the grade 10 HSPE also has a relatively high coefficient alpha of 0.88 for years 2011 and 2012 respectively. All of these coefficients suggest the construct is being measured consistently. All scores are at the high end of good reliability.

Validity

"Validity of an instrument means that it measures what it is designed to measure" (Spector, 1981, p. 14). Validity is the degree to which the assessment elicits the conceptual understanding and skills intended to be measured. Validity is a judgment about the relationships between a test score and its context (including the instructional practices and the examinee), the knowledge and skills it represents, the intended interpretations and uses, and the consequences of its interpretation and use (Educational Testing Service [ETS], 2010). That is, to what degree do test scores represent the intended construct?

Messick (1989) identified three strategies to establish the validity of the WASL and MSP for grade 8 and the HSPE for grade 10.

- Examine the content of the test in relation to the content of the domain of reference.
- Examine and probe the ways in which individuals respond to the items or tasks.
- Examine the relationships among responses to the tasks, items, or parts of the test; that is, the internal structure of test responses.

Messick's other three criteria concurrent, predictive, and consequential validity evidence are not relevant to the intended uses of the criterion-referenced WASL and MSP tests.

Content validity. The relationship between a test's content and the construct that the test was designed to measure can provide important evidence of validity. The construct of interest is operationally defined by state content standards and the test blueprints. The standards and test design specify the content, format, and scoring of items that are adequate measures of the knowledge and skills described in the content standards. Evidence that the items meet these specifications and represent the desired compliment of knowledge and skills, referenced by the

standards, supports the inference that students' scores on these items can appropriately be regarded as measures of the intended construct.

Logical analyses of test content in which experts judge the adequacy with which the test content conforms to the test specifications and represents the intended domain of content is evidence of validity. These reviews by experts can also be used to determine whether the test content contains material that is not relevant to the construct of interest. It was regular practice to have committees of teachers, content area experts, and professional test developers provide ongoing review, verification, and confirmation to ensure that the test content was aligned with the state standards (ETS, 2010). Intercorrelational analysis of the reading strands showed correlations between 0.62 and 0.74

Construct validity. Like the WASL and MSP, the grade 10 HSPE was examined for evidence based on test content, which includes a description of the Washington State standards, specifications and blueprints, item development process, item review process, the form construction process, and an alignment study. Relations to other variables also were examined to determine the relationships between test scores and measures of other variables external to the test. Correlations were examined between the Iowa Test of Educational Development (ITED) and other content areas (i.e., math and science). The degree to which the content area strand scores correlate, provides evidence of validity. In addition, evidence based on internal structure was considered. According to the 2011 OSPI report, "evidence of validity can be obtained from studies of the properties of the item scores and the relationship between these scores and scores on components of the test." To the extent that the score properties and relationships found are consistent with the definition of the construct measured by the test, support is gained for interpreting these scores as measures of the construct" (p. 115). Other validity measures such as

classical statistics (i.e., point-biserial correlations) and the IRT model-data fit analyses can be found in the 2009 OSPI technical report. A confirmatory factor analysis was done to examine construct validity. Also employed were the comparative fit index (CFI), root mean square error of approximation (RMSEA), chi-square (2), and the chi-square statistic divided by its associated degrees of freedom (df). The results of these analyses provide evidence of validity based on test content and content area constructs.

Internal validity. According to Jimenez-Buedo, and Miller (2010), internal validity is ensuring "that the treatment is isolated from potential confounds in order to make certain that the observed effect is attributable to the treatment" (p. 302). Gay et al., (2009) referred to internal validity as "the degree to which experimental results are attributable to the independent variable and not to another rival explanation" (p. 242). Internal validity is threatened when any event or condition unrelated to the treatment occurs during the study, which may affect the dependent variable (Campbell & Stanley, 1963; Gay et al., 2009).

Potential threats to validity

History. District records during the cohort periods outlined in this study were reviewed and no major events were found that required a significant or unequal interruption or postponement of schooling or testing. There were days where school did not meet as usual due to snow (i.e., inclement weather). Since these schools were in the same district, these days were identical for each school and the missed days were subsequently made up. Therefore, history is not a threat to the internal validity of this study.

Maturation. Maturation refers to changes (i.e., physical, intellectual, emotional) that occur naturally within individuals over a period of time (Campbell and Stanley, 1963; Gay et al., 2009). According to Gay et al. (2009), "maturation is more likely to be a problem in a study

designed to test the effectiveness of a psychomotor training program on three- year olds than in a study to compare two methods of teaching algebra' (p. 244). This study did follow cohorts of students for three years but is more comparable to Gay et al.'s algebra example. Therefore, maturation is not a significant risk to this study's internal validity.

Testing. Testing refers to the effect that taking a test has on scores of a subsequent test (Campbell & Stanley, 1963). Gay et al. (2009) referred to testing as *pretest sensitization* for it is the threat that a pre-test may affect the performance on a post-test. The general idea is that students who take a similar test for the second time do better even absent some form of treatment. The effects of testing are more likely when the two tests are taken in short time intervals and in studies that measure factual recall (Gay et al., 2009). Even though the 8th grade WASL/MSP and 10th grade HSPE have the same testing format, the length of time between the tests and the non-recall nature of each test make testing an unlikely threat to internal validity.

Instrumentation. The threat of instrumentation refers to unreliability of the measurement instruments, which may lead to uncertainty in the validity of the instrument's findings (Gay et al., 2009). According to Campbell and Stanley (1963), instrumentation is controlled when student responses are recorded by a fixed instrument such as a paper and pencil assessment. The Washington State assessments were developed specifically to measure student progress toward meeting the EALRs, GLEs, and PEs. The WASL, MSP, and HSPE combine multiple choice, short answer, and completion; therefore, the examinee's performance may differ decidedly from item to item. This heterogeneity of items in the reading assessments may tend to underestimate the reliability of scores as estimated by Cronbach's coefficient alpha. Even with this item heterogeneity, the WASL has alpha coefficients for reading in grade 8 of 0.88 for both the 2009 WASL and the 2010 MSP. The 2011 and 2012 coefficient alphas for the grade 10

HSPE each have relatively high coefficient alphas of 0.88. All scores are at the high end of good reliability, meaning the construct is being measured consistently. Instrumentation in this study is not likely a threat to its internal validity. As Campbell and Stanley (1963) stated, the design of this study—a non-experimental, associational explanatory design—controls the effects of history, maturation, testing, and instrumentation.

Statistical regression. Regression can be a major internal validity problem for non-experimental control group designs (Campbell & Stanley, 1963). The effects of regression can be mediated by ensuring that the two comparison groups have similar means with respect to initial assessment scores; "the more this similarity is confirmed by the scores on the pretest, the more effective this control becomes" (Campbell & Stanley, 1963, p. 48). The quasi-experimental design of this study—having groups that are similar in pretest scores—helps to control for this threat.

Differential selection of participants. The threat to internal validity is greatest when groups who were already formed are compared (Campbell and Stanley, 1963; Gay et al., 2009). In this study, students already were enrolled in one of the two high schools. Student attendance at each of the schools was determined by their place of residence. This study being an *ex post facto* design could not control for student school or group placement. The students included in this study (both schools) did take the WASL or MSP in the spring of their 8th grade year. The 8th grade assessment served as a pretest, which according to Spector (1981) is perhaps the most important variable for matching students in an *ex post facto* two-group design. Gay et al. (2009) supported this view when writing, if existing groups are included in the study, a pretest should be given to check for initial equivalence.

Grunwald and Mayhew (2008) posited that carrying out randomized experiments in education is not often plausible. Luellen, Shadish, and Clark (2005) go a step further when they say quasi-experimental research design, when done properly, may better reflect the complexity of the educational context in comparison to experimental design. It is acknowledged, however, that students in the comparison groups may have had undetected or uncontrolled differences that could potentially affect post-test scores.

Mortality. Mortality refers to a reduction in the numbers of research participants of one or more of the studied groups (Gay et al., 2009). This researcher placed no additional demands on students in the experimental group. Students from both schools (groups) were taught the district adopted honors English curriculum. Mortality in cases where there are no additional demands on any one group is rare (Gay et al., 2009).

Interactive effects. The most common interactive effect is selection-maturation interaction (Gay et al., 2009). The differences in student maturation may be due to differences in the qualities of the teachers. These teacher differences may then cause differences in post-test scores as a measure of the dependent variable. Gay et al. (2009) stressed the importance of the researcher controlling for these potential issues. Since this study examined the effects of ability grouping on student achievement on measures of reading or writing achievement, the fact that the researcher was unable to control for teacher variables is a limitation of this study.

External validity. External validity is the ability to extend the findings of research beyond the current study and sample (Campbell & Stanley, 1963; Gay et al., 2009; Spector, 1981). There are several threats to external validity that must be considered in establishing sound research design.

Population validity and ecological validity. Due to the *ex post facto* nature of this quasi-experimental study, many of the population and ecological validity threats are minimized. Data were collected in an unobtrusive manner meaning there was no interaction between researcher and student. This research design mediates many of the threats to external validity as described by Gay et al. (2009). These include pretest-treatment effects, treatment diffusion, experimenter effects, and reactive arrangements. All of these threats are dependent on the researcher having contact or influence with the participants. In an *ex post facto* design such as this, contact did not occur.

Specificity of variables. A final threat to external validity identified by Gay et al. (2009) pertaining to this study is specificity of variables. Gay et al. (2009) cited this threat when "the study is conducted with a specific kind of participant, using specific measuring instruments, at a specific time, and under special circumstances" (p. 247). Students in this study were assessed on the same day and at the same time for each of the identified cohorts. In addition, multiple post-treatment assessments were given, which according to Gay et al. (2009) "is the only way to assess the generalizability of findings over time" (p. 248). The interaction of history and treatment effects, which Gay et al. (2009) described as extraneous events that could alter the results of the study were also considered. The researcher was present during the entire period of the study and is not aware of any disruptive events that would have affected a significant number of group members.

Population validity. Threats to population validity often occur at the data analysis stage because researchers fail to disaggregate their data, incorrectly assuming that their findings are invariant across all subsamples inherent in their study" (Onwuegbuzie, 2003, p. 84). To account for threats to population validity, this study looked at two cohorts from two diverse high schools

spanning a three-year period. Data were disaggregated along the lines of Ethnicity, Gender, SES, and ELL. Resources such as time and money were limited so the study may be considered by many to be relatively small, which poses a threat to external validity. Onwuegbuzie (2003) stated that the distribution of scores was sample specific so generalizing their meaning outside the sample is questionable.

Summary of threats to internal and external validity

"Providing information about sources of invalidity and rival explanations (a) allows readers to better contextualize the underlying findings, (b) promotes external replications, (c) provides a direction for future research, and (d) advances the conducting of validity meta analyses and thematic effect sizes" (Onwuegbuzie, 2003, p. 87). The *ex post facto*, non-experimental, explanatory associational design of this study facilitates confidence in the internal and external validity of the study. According to Jimenez-Buedo, and Miller (2005):

> internal validity and external validity stand in a relationship best described as a trade-off: the more we ensure that the treatment is isolated from potential confounds in order to make certain that the observed effect is attributable to the treatment, the more unlikely it is that the experimental results can be representative of phenomena of the outside world. (p. 302)

Of course, no research study is methodologically pure; including this one. As identified, there are threats to both internal and external validity. The research design is methodologically sound, however, as many threats to its validity have been accounted for or marginalized.

Methodology summary

The purpose of this study is to determine the extent to which student ability grouping (homogeneous versus heterogeneous) affects high ability students' academic achievement. Two

cohorts of students from two different high schools were selected for study. Pre-achievement for each of the groups was measured utilizing the 8th grade Washington State reading assessments. The study design was quasi-experimental and *ex post facto* so randomization of student participants was not possible. Each school served as a proxy for ability grouping; with School 0 grouping honors students homogeneously and School 1 grouping honors students with non-honors students (heterogeneously). To control for school effect, a chi-square analysis was completed. Student traits of Gender, Ethnicity, SES, and Special education status, and number of student Absences were compared across schools. A multiple regression then was used to identify the best set of predictor variables. Tests for co-linearity were also run to solidify the conclusions regarding the effects of each of the independent variables. The dependent variable of academic achievement was measured by the 10th grade Washington State Assessment in reading (HSPE) and student scores on the reading and writing sections of the PSAT. Data, therefore, were analyzed utilizing chi-square and multiple regression using SPSS version 20. Significance was set at the .05 probability level or higher. Potential threats to the internal and external validity of this study at the research design/data collection and data analysis level were identified. One of the goals of this study is to allow for replication. Gerring (2011) stated "to facilitate replication, a research design must be conducted in such a way that future scholars can reproduce its results...replicability is simply a method of checking a study's internal validity" (p. 628). Chapter IV is a presentation of the results and statistical analysis of this study.

Chapter IV

Analysis of data

The purpose of this study was to determine if there is a measurable academic benefit to homogeneously grouping high school honors English students in a diverse, suburban school district in Washington State. Specifically, the effect of the predictor variable—type of grouping (mixed-ability/like-ability)—on the dependent variable of student achievement was analyzed while controlling for student variables associated with student achievement. A standardized 8th grade state assessment in reading served as a pre-treatment measure and the standardized 10th grade state assessment in reading served as a post-treatment measure. In addition, the PSAT critical reading and critical writing scores of students were used as outcome variables. By concentrating on the variable of student grouping and analyzing quantitative data collected before, during, and after student exposure to the ability grouping variable, the objective of this study was to produce research-based evidence to assist policymakers, educators, and parents in their decisions on whether to group students by ability. Furthermore, the goal is to have data that assist policymakers and practitioners in the development of structures that will maximize the learning and achievement of all students.

Research questions

Specific and intentional SPSS version 20 outputs will be used to answer the following research questions:

- To what extent, if any, does ability grouping of high achieving students defined as 9th and 10th grade honors English students in a suburban Washington State school district, affect their performance on state and pre-college assessments of reading and writing achievement when controlling for student mutable variables?

Subsidiary research questions:

- To what extent, if any, does placement of 2009–2011 honors English students in homogeneous ability grouped English classes have on their subsequent performance on state and pre-college assessments of reading and writing achievement?
- To what extent, if any, does placement of 2009–2011 honors English students in heterogeneous ability groups have on their subsequent performance on state and pre-college assessments of reading and writing achievement?
- To what extent, if any, does placement of 2010–2012 honors English students in homogeneous ability grouped English classes have on their subsequent performance on state and pre-college assessments of reading and writing achievement?
- To what extent, if any, does placement of 2010–2012 honors English students in heterogeneous ability groups have on their subsequent performance on state and pre-college assessments of reading and writing achievement?

Description of the sample

Cohort 0. Cohort 0 represents students who were in the 8th grade during the 2008–2009 school year. Each student completed the state assessment in reading during the spring of 2009. The students then experienced honors English instruction during the 2009 –2010 and 2010–2011 school years—their 9th and 10th grade years. This honors English instruction was delivered either in homogeneous ability groups (School 0) or heterogeneous ability groups (School 1).

School 0 had the following student demographic characteristics: (a) 0 % ELL; (b) 11.5% free and/or reduced lunch (FRL); (c) 57.6 % male; (d) 55.6 % white, 31.7 % Asian, 3.7 % black, 3.3% Hispanic, 5.8% Other or multiracial; 0.8% Special education. School 1 had (a) 0% ELL,

(b) 13.2% FRL; (c) 56.6% male; (d) 57.2% white, 27.6% Asian, 3.5 % black, 4.1% Hispanic, 7.6% Other or multiracial; 0.7% Special education.

Cohort 1. Cohort 1 represents students who were in the 8th grade during the 2009–2010 school year. Each student completed the state assessment in reading during the spring of 2010. The students then experienced honors English instruction during the 2010 –2011 and 2011–2012 school years—their 9th and 10th grade years. This honors English instruction was delivered either in homogeneous ability groups (School 0) or heterogeneous ability groups (School 1).

School 0 had the following student demographic characteristics: (a) 0 % ELL; (b) 24.4% free and/or reduced lunch (FRL); (c) 51.9 % male; (d) 51.5 % white, 27.1 % Asian, 6.8 % black, 6.8% Hispanic, 7.8% Other or multiracial; 0.3% Special education. School 1 had (a) 0% ELL, (b) 21.1% FRL; (c) 60.0% male; (d) 60.0% white, 23.9% Asian, 6.7 % black, 2.8% Hispanic, 6.7% Other or multiracial; 0.6% Special education.

Data for both cohorts were acquired with the assistance of the school district's assessment office utilizing the district's database software. All proper protocols as required by the Seton Hall University IRB were followed and maintained throughout the duration of this study to ensure student anonymity and confidentiality.

General results

Chi square analysis. A chi-square analysis for each cohort (see Appendix A) was completed independently to determine if there were any significant differences between the two groups (Schools) in regard to the independent student variables of (a) ELL, (b) FRL, (c) Gender, (d) Ethnicity, and (e) Special education.

Cohort 0. In regard to each of the independent variables, there were no significant differences between the two groups. (a) ELL: χ^2 = constant, (b) FRL: $\chi^2 = .213$, df = 1, $p \leq .644$,

(c) Gender: $\chi^2 = .042$, df = 1, $p \leq .838$ (d) Ethnicity: $\chi^2 = 4.038$, df = 5, $p \leq .544$, (e) Special education: $\chi^2 = .021$, df = 1, $p \leq .885$.

Cohort 1. In regard to each of the independent variables, there were no significant differences between the two groups. (a) ELL: χ^2 = constant, (b) FRL: $\chi^2 = .682$, df = 1, $p \leq .409$, (c) Gender: $\chi^2 = 2.989$, df = 1, $p \leq .084$ (d) Ethnicity: $\chi^2 = 5.516$, df = 5, $p \leq .356$, (e) Special education: $\chi^2 = .125$, df = 1, $p \leq .724$.

Independent samples *t*-test

An independent samples *t*-test (see Appendix B) was completed on the independent student variables of full day absences and grade 8 state reading assessment to determine if there were any significant differences between the two groups (schools). The grade 8 state reading assessment for Cohort 0 was the WASL and for Cohort 1 was the MSP. The results for both independent variables fail to reject the null hypothesis of no difference between schools. Full-day absences Cohort 0: $t = -.764$, df =381, $p \leq .445$. Full-day absences Cohort 1: $t = -.840$, df = 473, $p \leq .401$. WASL Cohort 0: $t = 1.906$, df = 338, $p \leq .058$. MSP Cohort 1: $t = -1.812$, df = 439, $p \leq .071$.

Descriptive statistics

Descriptive statistics were determined for each of the two cohorts. Table 5 is a presentation of the statistics for Cohort 0 and Table 6 the statistics for Cohort 1.

Table 5

Descriptive statistics Cohort 0

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
Num_of_full_day_absences	383	0	71	12.66	12.143	1.497	.125
Grade_8_WASL_Reading_2009	375	369	500	426.98	19.984	.949	.126
Grade_10_HSPE_Reading_2011_	384	390	525	459.27	26.824	.421	.125
PSAT_Reading_2010	38	38	69	52.05	7.472	.467	.383
PSAT_Writing_2010	38	36	71	50.71	8.137	.539	.383
PSAT_Reading_2011	321	18	80	52.63	9.647	.125	.136
PSAT_Writing_2011	321	26	80	49.22	8.651	.542	.136
Valid N (listwise)	32						

Table 6

Descriptive statistics Cohort 1

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
Num_of_full_day_absences	475	0	124	13.65	12.057	2.557	.112
Grade_8_MSP_Reading_2010	441	348	500	423.15	20.673	.332	.116
Grade_10_HSPE_Reading_2012	469	390	525	446.05	25.243	.674	.113
PSAT_Reading_2011	453	23	73	48.17	8.062	-.116	.115
PSAT_Writing_2011	453	20	71	44.44	8.039	.135	.115
PSAT_Reading_2012	235	31	80	52.86	9.265	.100	.159
PSAT_Writing_2012	235	31	72	50.96	9.123	.033	.159
Valid N (listwise)	205						

The dependent variables on which of the regression models were completed for Cohort 0 are: Grade 10 HSPE Reading 2011, PSAT Reading 2010, PSAT Writing 2010, PSAT Reading 2011, and PSAT Writing 2011. The dependent variables for Cohort 1 are: Grade 10 HSPE Reading 2012, PSAT Reading 2011, PSAT Writing 2011, PSAT Reading 2012, and PSAT Writing 2012. Field (2009) stated the minimum sample size for a statistically meaningful multiple regression is 50 + 8k, where k is the number of predictors. The predictor variables for both cohorts are School (proxy for type of grouping), FRL, Gender, Ethnicity (Black/not Black,

Asian/not Asian, Hispanic/not Hispanic, Other: multiracial/not other: multiracial), Special education status, full day Absences, and Grade 8 assessment scores. Considering the values found in Tables 5 and 6 above, the sample size (n values) for the regressions of Cohort 0 and Cohort 1 exceed the number called for by (Field, 2009). Two of the models for Cohort 0 with the dependent variables of PSAT Reading 2010 and PSAT Writing 2010 were the exception. These two models only had a sample size (n) of 38, significantly less than the number supported by Field (2009).

Skewness was included in the analysis to determine the appropriateness of using common inferential statistical measures. Skewness of less than plus or minus 1.0 is generally thought to indicate that the variable being studied is at least approximately normal (Leech, Barrett, & Morgan, 2011). All variables, with the exception of number of full day Absences, are within the range of normal distribution. This is true for both cohorts. The full day absences skewness for Cohort 0 = 1.497 and for Cohort 1 = 2.557. Considering the special population of students; that is, the students in this study are all honors students, this skewness is not unexpected. Leech, Barrett, and Morgan (2011) suggested that if the variable is markedly skewed, then it is prudent to "either transform the data or use a nonparametric statistic" (p. 22). Therefore, the data for the variable number of full day Absences were transformed into a dichotomous variable. Using the median value of 10 absences (over two years) as the midpoint, absence values below (n = 431) were coded 0 and values above (n = 432) were coded 1. This leaves all remaining interval variables within the range of normal distribution.

Correlation analysis

A correlation analysis was completed to identify the relationship between the dependent variables and the independent variables for each cohort respectively (see Appendix C). The

correlation coefficient represents the linear relationship between two variables. Values of correlation are always between − 1 and + 1. A correlation coefficient value of + 1 indicates a perfectly positive linear correlation, while a value of −1 indicates a perfectly negative linear correlation. A correlation of 0 indicates no linear relationship between variables.

A general examination of the correlation table of Cohort 0 indicates moderate to strong correlations between the dependent variables. This is expected as they all measure a student's achievement in either reading or writing. The only significant correlations between predictors are Gender and Absences (Pearson's r = .164) and Special education and Absences (Pearson's r = .222). These are weak correlations and not indicative of any possible multicollinearity issues. Of particular note, the variable School (proxy for type of grouping) was not significantly correlated with any of the dependent variables.

As does Cohort 0, Cohort 1 shows moderate to strong significant correlations between the dependent variables. These correlations range from (Pearson's r = .345 to Pearson's r = .787). There are very weak significant correlations between FRL and Ethnicity (Pearson's r = .115), Gender and Absences (Pearson's r = .108), and Gender and Grade 8 MSP Reading (Pearson's r = .198). This indicates a possible absence of multicollinearity concerns between the independent variables of Cohort 1. Therefore, there is little concern for the suppression of independent variables when running multiple regressions on either of these cohorts. The variable school was significantly, but very weakly, correlated with the independent variables Gender (Pearson's r = .101) and Grade 8 MSP Reading (Pearson's r = .198). The variable school was also weakly and significantly correlated with each of the dependent variables (ranging from Pearson's r = .176 to Pearson's r = .265). Of particular note, the weak correlations were all positive, indicating scores on the dependent variable are positively correlated with the school coded as 1.

Results: Cohort 0

Cohort 0 represents students who were in the 8th grade during the 2008–2009 school year. Each student completed the state assessment in reading during the spring of 2009. The students then experienced honors English instruction during the 2009–2010 and 2010–2011 school years—their 9th and 10th grade years. Either honors English instruction was delivered in homogeneous ability groups (School 0) or heterogeneous ability groups (School 1). Some of the students in this cohort took the PSAT in the fall of their sophomore year (2010). Scores for the reading and writing portion of this PSAT assessment were collected. Many students completed the PSAT in the fall of their junior year (2011). Scores for the reading and writing portion of the PSAT assessment were collected. Students also took the High School Proficiency Exam (HSPE) in the spring of their sophomore year (2011). Student scores for the reading portion of this assessment were collected.

Using SPSS version 20 and the linear regression analysis, models for each of the dependent or grouping variables were analyzed. ELL will not be entered in any of the following models as it has been previously identified as a constant. ELL is a constant because no ELL students were enrolled in honors English in either of the schools studied. Tables 7–10 show the results of the regression analysis on the dependent/outcome variable HSPE Reading 2011 when the 2010 Grade 8 WASL reading scores, Student absences, Ethnicity (A-NA, B-NB, H-NH, OM-NOM), Special education, FRL, School, and Gender were used in the model as independent/predictor variables. In terms of analyzing the influence of ethnicity, it should be noted that for all the following models white serves as the reference category. This means that if the computed β value is negative, white is the race identified as having a more significant influence on the model.

Tables 7–10 show the results of the regression analysis on the dependent/outcome variable HSPE Reading 2011 when the 2010 Grade 8 WASL reading scores, Student absences, Ethnicity (A-NA, B-NB, H-NH, OM-NOM), Special education, FRL, School, and Gender were used in the model as independent/predictor variables.

Table 7

Multiple regression variables entered HSPE reading 2011

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Grade_8_WASL_Reading_2009, Absences_by_Median, FRL, B_NB, H_NH, Sped, OM_NOM, Gender, School, A_NA[b]		Enter

a. Dependent Variable: Grade_10_HSPE_Reading_2011_

b. All requested variables entered.

Table 8

Model summary mulitple regression HSPE reading 2011

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.403[a]	.162	.139	25.125

a. Predictors: (Constant), Grade_8_WASL_Reading_2009, Absences_by_Median, FRL, B_NB, H_NH, Sped, OM_NOM, Gender, School, A_NA

The adjusted R^2 for this model indicates that 13.9% of the variance in student performance on the 2011 HSPE reading assessment can be explained by school (proxy for type of grouping), FRL, Gender, Ethnicity (Black/not Black, Asian/not Asian, Hispanic/not Hispanic, Other: multiracial/not other: multiracial), Special education, Absences, and Grade 8 WASL reading.

Table 9

Multiple regression ANOVA HSPE reading 2011

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	43599.813	10	4359.981	6.907	.000[b]
	Residual	225358.871	357	631.257		
	Total	268958.685	367			

a. Dependent Variable: Grade_10_HSPE_Reading_2011_

b. Predictors: (Constant), Grade_8_WASL_Reading_2009, Absences_by_Median, FRL, B_NB, H_NH, Sped, OM_NOM, Gender, School, A_NA

The ANOVA table indicates the model is statistically significant $F = 6.907$, $df = 10,357$, $p < .001$.

Table 10

Multiple regression coefficients table HSPE reading 2011

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	253.046	29.035		8.715	.000					
	School	5.169	2.757	.093	1.875	.062	.034	.099	.091	.963	1.039
	FRL	-10.142	4.109	-.123	-2.468	.014	-.125	-.130	-.120	.947	1.056
	Gender	-5.521	2.680	-.101	-2.060	.040	-.097	-.108	-.100	.976	1.025
	B_NB	-4.748	7.480	-.031	-.635	.526	-.041	-.034	-.031	.972	1.029
	A_NA	-3.629	3.098	-.061	-1.171	.242	-.005	-.062	-.057	.853	1.173
	H_NH	3.135	7.224	.021	.434	.665	.024	.023	.021	.964	1.037
	OM_NOM	-6.173	5.433	-.056	-1.136	.257	-.069	-.060	-.055	.953	1.049
	Sped	-9.949	17.946	-.027	-.554	.580	-.044	-.029	-.027	.985	1.015
	Absences_by_Median	-1.972	2.705	-.036	-.729	.466	-.040	-.039	-.035	.942	1.062
	Grade_8_WASL_Reading_2009	.495	.068	.365	7.292	.000	.342	.360	.353	.937	1.067

a. Dependent Variable: Grade_10_HSPE_Reading_2011_

Examination of the coefficients table indicates that there are three significant predictors: FRL $\beta = -.123$, $t = -2.458$, $p \leq .014$; Gender $\beta = -101$, $t = -2.060$, $p \leq .040$; Grade 8 WASL reading $\beta = .365$, $t = 7.292$, $p < .001$. The negative β value for both FRL and Gender signifies that students who are not FRL and those who are male (not female) are favored. While the positive β for Grade 8 WASL reading favors those students with higher test scores. Squaring the β values of each significant predictor indicates that FRL explains 1.5% of the model's variance; Gender 1.0%, and Grade 8 WASL reading 13.3%. School is not a significant predictor. Therefore, type of grouping did not significantly contribute to this model. An analysis to determine the presence

of multicollinearity issues is to examine both the tolerance and VIF values. A tolerance value < $(1 - R^2)$ may indicate a multicollinearity problem (Leech, Barrett, & Morgan, 2011). The predictor Asian/not Asian is .853, which is less than .861 $(1 - R^2)$. However, since the VIF statistic (1.173) for Asian/not Asian does not exceed the value of 2, multicollinearity is most likely not an issue.

Tables 11–13 show the results of the regression analysis on the dependent/outcome variable PSAT Reading 2010 when the 2010 Grade 8 WASL reading scores, Student absences, Ethnicity (A-NA, B-NB, H-NH, OM-NOM), Special education, FRL, School, and Gender were used in the model as independent/predictor variables.

Table 11

Multiple regression variables entered PSAT reading 2010

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Grade_8_WASL_Reading_2009, B_NB, OM_NOM, H_NH, A_NA, School, FRL, Absences_by_Median, Gender[b]		Enter

a. Dependent Variable: PSAT_Reading_2010

b. All requested variables entered.

Note: The variable of special education has been removed due to it being a constant. This means no students in this model were classified as special education.

Table 12

Multiple regression model summary PSAT reading 2010

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.640[a]	.410	.213	6.714

a. Predictors: (Constant), Grade_8_WASL_Reading_2009, B_NB, OM_NOM, H_NH, A_NA, School, FRL, Absences_by_Median, Gender

Table 13

Multiple regression ANOVA table PSAT reading 2010

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	844.529	9	93.837	2.082	.068[b]
	Residual	1217.038	27	45.075		
	Total	2061.568	36			

a. Dependent Variable: PSAT_Reading_2010

b. Predictors: (Constant), Grade_8_WASL_Reading_2009, B_NB, OM_NOM, H_NH, A_NA, School, FRL, Absences_by_Median, Gender

This is not a significant model ($p \leq .068 > .05$). No further consideration is warranted.

Tables 14–17 show the results of the regression analysis on the dependent/outcome variable PSAT Writing 2010 when the 2010 Grade 8 WASL reading scores, Student absences, Ethnicity (A-NA, B-NB, H-NH, OM-NOM), Special education, FRL, School, and Gender were used in the model as independent/predictor variables.

Table 14

Multiple regression variables entered PSAT writing 2010

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Grade_8_WASL_Reading_2009, B_NB, OM_NOM, H_NH, A_NA, School, FRL, Absences_by_Median, Gender[b]		Enter

a. Dependent Variable: PSAT_Writing_2010

b. All requested variables entered.

Note: The variable of special education has been removed due to it being a constant. This means no students in this model were classified as special education.

Table 15

Multiple regression model summary for PSAT writing 2010

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.770[a]	.593	.457	6.078

a. Predictors: (Constant), Grade_8_WASL_Reading_2009, B_NB, OM_NOM, H_NH, A_NA, School, FRL, Absences_by_Median, Gender

The adjusted R^2 for this model indicates that 45.7% of the variance in student performance on the 2010 PSAT Writing assessment can be explained by school (proxy for type of grouping), FRL, Gender, Ethnicity (Black/not Black, Asian/not Asian, Hispanic/not Hispanic, Other: multiracial/not other: multiracial, Absences, and Grade 8 WASL reading.

Table 16

Multiple regression ANOVA table PSAT writing 2010

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1452.240	9	161.360	4.368	.001[b]
	Residual	997.490	27	36.944		
	Total	2449.730	36			

a. Dependent Variable: PSAT_Writing_2010

b. Predictors: (Constant), Grade_8_WASL_Reading_2009, B_NB, OM_NOM, H_NH, A_NA, School, FRL, Absences_by_Median, Gender

The ANOVA table indicates the model is statistically significant $F = 4.368$, $df = 9,27$, $p \leq .001$.

Table 17

Multiple regression coefficients table PSAT writing 2010

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-19.005	23.348		-.814	.423					
	School	-5.014	5.039	-.139	-.995	.329	.123	-.188	-.122	.769	1.300
	FRL	1.382	3.639	.053	.380	.707	-.148	.073	.047	.782	1.279
	Gender	-.534	2.461	-.031	-.217	.830	-.229	-.042	-.027	.724	1.382
	B_NB	18.784	6.434	.374	2.919	.007	.375	.490	.359	.917	1.090
	A_NA	-8.790	3.188	-.423	-2.757	.010	-.423	-.469	-.339	.640	1.561
	H_NH	-5.657	5.144	-.157	-1.100	.281	-.212	-.207	-.135	.738	1.355
	OM_NOM	-4.606	3.535	-.176	-1.303	.204	-.073	-.243	-.160	.829	1.207
	Absences_by_Median	-5.916	2.369	-.360	-2.498	.019	-.313	-.433	-.307	.725	1.379
	Grade_8_WASL_Reading_2009	.187	.056	.435	3.315	.003	.304	.538	.407	.875	1.143

a. Dependent Variable: PSAT_Writing_2010

Examination of the coefficients table indicates that there are four significant predictors: Black/not Black $\beta = .374$, $t = 2.919$, $p \leq .007$; Asian/not Asian $\beta = -.423$, $t = -2.757$, $p \leq .010$; Absences $\beta = -.360$, $t = -2.498$, $p \leq .019$; and Grade 8 WASL reading $\beta = .435$, $t = 3.315$, $p \leq .003$. The positive β value for Black/not Black and Grade 8 WASL reading signifies that students who are black and those who scored higher on the Grade 8 WASL are favored. The negative β value for Asian/not Asian and attendance favor students who are white and those who had fewer absences. Squaring the β values of each significant predictor indicates that Black/not Black explains 14% of the model's variance; Asian/not Asian 17.9%, attendance 13.0%, and Grade 8 WASL reading 18.9%. Therefore, Grade 8 WASL reading is the strongest predictor for this model. The variable school (proxy for type of grouping) is not a significant predictor. An analysis of the model's tolerance and VIF values indicates no apparent issues with multicollinearity.

Tables 18–21 show the results of the regression analysis on the dependent/outcome variable PSAT Reading 2011 when the 2010 Grade 8 WASL reading scores, student absences, Ethnicity (A-NA, B-NB, H-NH, OM-NOM), Special education, FRL, School, and Gender were used in the model as independent/predictor variables..

Table 18

Multiple regression variables entered PSAT reading 2011

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Grade_8_WASL_Reading_2009, H_NH, B_NB, Absences_by_Median, FRL, Sped, Gender, OM_NOM, School, A_NA[b]		Enter

a. Dependent Variable: PSAT_Reading_2011

b. All requested variables entered.

Table 19

Multiple regression model summary PSAT reading 2011

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.459[a]	.210	.184	8.768

a. Predictors: (Constant), Grade_8_WASL_Reading_2009, H_NH, B_NB, Absences_by_Median, FRL, Sped, Gender, OM_NOM, School, A_NA

The adjusted R^2 for this model indicates that 18.4% of the variance in student performance on the 2011 PSAT Reading assessment can be explained by school (proxy for type of grouping), FRL, Gender, Ethnicity (Black/not Black, Asian/not Asian, Hispanic/not Hispanic, Other: multiracial/not other: multiracial), Special education, Absences, and Grade 8 WASL reading.

Table 20

Multiple regression ANOVA PSAT reading 2011

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6082.531	10	608.253	7.911	.000[b]
	Residual	22834.742	297	76.885		
	Total	28917.273	307			

a. Dependent Variable: PSAT_Reading_2011

b. Predictors: (Constant), Grade_8_WASL_Reading_2009, H_NH, B_NB, Absences_by_Median, FRL, Sped, Gender, OM_NOM, School, A_NA

The ANOVA table indicates the model is statistically significant $F = 7.911$, $df = 10, 297$, $p < .001$.

Table 21

Multiple regression coefficients table PSAT reading 2011

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-26.503	11.033		-2.402	.017					
	School	3.507	1.119	.166	3.133	.002	.087	.179	.162	.951	1.051
	FRL	-3.549	1.723	-.110	-2.060	.040	-.093	-.119	-.106	.929	1.076
	Gender	-2.340	1.023	-.119	-2.288	.023	-.120	-.132	-.118	.979	1.022
	B_NB	-4.343	3.002	-.075	-1.446	.149	-.079	-.084	-.075	.976	1.024
	A_NA	-1.453	1.164	-.070	-1.249	.213	-.020	-.072	-.064	.854	1.170
	H_NH	1.493	3.182	.025	.469	.639	.025	.027	.024	.974	1.026
	OM_NOM	.158	2.002	.004	.079	.937	-.012	.005	.004	.939	1.065
	Sped	1.399	6.284	.012	.223	.824	-.022	.013	.011	.980	1.021
	Absences_by_M	-3.640	1.036	-.187	-3.513	.001	-.157	-.200	-.181	.939	1.065
	Grade_8_WASL_Reading_2009	.191	.026	.394	7.422	.000	.340	.396	.383	.941	1.063

a. Dependent Variable: PSAT_Reading_2011

Examination of the coefficients table indicates that there are five significant predictors: School $\beta = .166$, $t = 3.133$, $p \leq .002$; FRL $\beta = -.110$, $t = 2.060$, $p \leq .040$; Gender $\beta = -.119$, $t = -2.288$, $p \leq .023$, Absences $\beta = -.187$, $t = -3.513$, $p \leq .001$ and Grade 8 WASL reading $\beta = .394$, $t = 7.422$, $p < .001$. The positive β value for school and Grade 8 WASL reading signifies that students attending School 1 (heterogeneous grouping) and those who scored higher on the Grade 8 WASL have a slightly positive influence on the dependent variable of PSAT Reading 2011. The

negative β value for FRL, Gender, and Absences indicates that students who were not FRL, who are males (not females), and those who had fewer absences have a slightly positive influence on the dependent variable. Squaring the β values of each significant predictor indicates that School explains 2.8% of the model's variance, FRL 1.2 %, Gender 1.4%, Attendance 3.5%, and Grade 8 WASL reading 15.5%. Therefore, Grade 8 WASL reading is the greatest predictor for this model. An analysis of the model's tolerance and VIF values indicates no apparent issues with multicollinearity.

Tables 22–25 show the results of the regression analysis on the dependent/outcome variable PSAT Writing 2011 when the 2010 Grade 8 WASL reading scores, Absences, Ethnicity (A-NA, B-NB, H-NH, OM-NOM), Special education, FRL, School, and Gender were used in the model as independent/predictor variables.

Table 22

Multiple regression variables entered PSAT writing 2011

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Grade_8_WASL_Reading_2009, H_NH, B_NB, Absences_by_Median, FRL, Sped, Gender, OM_NOM, School, A_NA[b]		Enter

a. Dependent Variable: PSAT_Writing_2011

b. All requested variables entered.

Table 23

Multiple regression model summary PSAT wrting 2011

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.483[a]	.233	.207	7.767

a. Predictors: (Constant), Grade_8_WASL_Reading_2009, H_NH, B_NB, Absences_by_Median, FRL, Sped, Gender, OM_NOM, School, A_NA

The adjusted R^2 for this model indicates that 20.7% of the variance in student performance on the 2011 PSAT Writing assessment can be explained by school (proxy for type of grouping), FRL, Gender, Ethnicity (Black/not Black, Asian/not Asian, Hispanic/not Hispanic, Other: multiracial/not other: multiracial), Special education, Absences, and Grade 8 WASL reading.

Table 24

Multiple regression ANOVA PSAT writing 2011

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5444.507	10	544.451	9.024	.000[b]
	Residual	17919.026	297	60.333		
	Total	23363.532	307			

a. Dependent Variable: PSAT_Writing_2011

b. Predictors: (Constant), Grade_8_WASL_Reading_2009, H_NH, B_NB, Absences_by_Median, FRL, Sped, Gender, OM_NOM, School, A_NA

The ANOVA table indicates the model is statistically significant $F = 9.024$, $df = 10, 297$, $p < .001$.

Table 25

Multiple regression coefficients table PSAT writing 2011

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-26.773	9.774		-2.739	.007					
	School	2.927	.991	.154	2.952	.003	.069	.169	.150	.951	1.051
	FRL	-3.508	1.526	-.121	-2.298	.022	-.099	-.132	-.117	.929	1.076
	Gender	-1.755	.906	-.100	-1.938	.054	-.093	-.112	-.098	.979	1.022
	B_NB	-5.825	2.660	-.113	-2.190	.029	-.110	-.126	-.111	.976	1.024
	A_NA	-1.572	1.031	-.084	-1.525	.128	-.011	-.088	-.077	.854	1.170
	H_NH	.587	2.819	.011	.208	.835	.020	.012	.011	.974	1.026
	OM_NOM	-2.892	1.773	-.086	-1.631	.104	-.094	-.094	-.083	.939	1.065
	Sped	-4.308	5.567	-.040	-.774	.440	-.068	-.045	-.039	.980	1.021
	Absences_by_M	-2.422	.918	-.138	-2.639	.009	-.117	-.151	-.134	.939	1.065
	Grade_8_WASL_Reading_2009	.183	.023	.421	8.031	.000	.382	.422	.408	.941	1.063

a. Dependent Variable: PSAT_Writing_2011

Examination of the coefficients table indicates that there are five significant predictors: School $\beta = .154$, $t = 2.952$, $p \leq .003$; FRL $\beta = -121$, $t = -2.298$, $p \leq .022$; Black/not Black $\beta = -.113$, $t = -2.190$, $p \leq .029$; Absences $\beta = -.138$, $t = -2.639$, $p \leq .009$; and Grade 8 WASL reading $\beta = .421$, $t = 8.031$, $p < .001$. School is a significant predictor with a positive β of .154. This indicates that School 1 (heterogeneous grouping) has a slightly positive influence on the dependent variable of PSAT Writing 2011. Grade 8 WASL also has a positive β signifying that higher student scores have positive influence on the dependent variable. The negative β for FRL, Black/not Black and Absences signify that students who are not FRL, white, and those with fewer absences have a positive effect on scores on the dependent variable of PSAT Writing 2011.

Squaring the β values of each significant predictor indicates that school explains 2.4% of the model's variance, FRL 1.5%, Black/not Black 1.3%, Attendance 1.9% and Grade 8 WASL reading 17.7%. Therefore, Grade 8 WASL is the strongest predictor for this model. An analysis of the model's tolerance and VIF values indicates no apparent issues with multicollinearity.

Summary of Results: Cohort 0

Students in Cohort 0 experienced honors English instruction for two consecutive years. Cohort 0 students were in the 8th grade during the 2008–2009 school year. Upon entering their first year of high school, the 9th grade, students chose and were given an honors English curriculum. In School 0 the curriculum was delivered in a homogeneous environment where all students in the honors English classroom were honors students. In School 1, the curriculum was delivered in a heterogeneous environment where approximately half the students were honors students and the other half were not (ie. core English students). Students continued in these like-ability (School 0) or mixed-ability (School 1) honors English classes through their 10th grade years. The results from five state and/or pre-college assessments in reading and/or writing were collected.

High school proficiency exam (HSPE) in reading 2011

This state standardized assessment was taken by students on the same day in March, 2011; the spring of the students' sophomore year. The ANOVA (see Table 9) shows the model was statistically significant ($p \leq .001$). The adjusted R^2 for this model indicates the model accounts for 13.9% of the variance in the dependent variable. The multiple regression coefficients table (see Table 10) shows that School (proxy for grouping type) was not a significant contributor to the HSPE Reading 2011 model ($p \leq .062$). The three significant predictors were FRL (1.5% variance), Gender (1.0%), and Grade 8 WASL reading scores (13.3%). The variables FRL and Gender each contributed negatively to the model. Meaning students who were FRL and males (not females) had a negative contribution to the model. The variable Grade 8 WASL reading assessment (positive β) positively contributed to the scores on the HSPE Reading 2011 assessment. Grade 8 WASL scores were the most significant predictor

at 13.3 % of the variance. This was nearly 8.9 times stronger than the next highest predictor (FRL). The fact that males outperformed females is a bit of a surprise. Females are generally thought to perform better on measures of reading achievement. Due to a tolerance value for the predictor, Asian/not Asian that is slightly lower than $(1 - R^2)$ a multicollinearity issue, although not likely, is possible.

PSAT Reading 2010

The PSAT in reading was administered to interested students in the fall of their sophomore year. The ANOVA (see Table 13) indicates that the multiple regression model with the dependent variable PSAT Reading was not a significant model ($p \leq .068$).

PSAT Writing 2010

The PSAT in writing was administered to interested students in the fall of their sophomore year. The fact that students could choose to take the PSAT is a limitation to this study. That is, the higher performing or motivated students may have been the ones choosing to take the PSAT. The ANOVA (see Table 16) indicates that the multiple regression model with the dependent variable PSAT Writing 2010 was a significant model ($p \leq .001$). The adjusted R^2 for this model indicates the model accounts for 45.7% of the variance in the dependent variable. The coefficient table (see Table 17) shows the variable School was not a significant predictor ($p \leq .329$). The four significant predictors were the variables: Black/not Black (favoring black), Asian/not Asian (favoring white), Absences (favoring fewer absences), and Grade 8 WASL reading (favoring higher scores). Grade 8 WASL reading was the strongest predictor (18.9%) followed by Asian/not Asian (17.9%), Black/not Black (14%) and absences (13%). Only 38 students in Cohort 0 took the PSAT Writing 2010. This small sample size does not meet the minimum sample size needed for multiple regressions (Field, 2009). Field (2009) established the

ideal minimum regression sample size to be 80 + 5k, (where k equals the number of predictors). Therefore, this particular model with the dependent variable of PSAT Writing 2010 does not warrant further consideration.

PSAT Reading 2011

The PSAT in reading was administered to interested students in October 2011 of their junior year. The ANOVA table (see Table 20) indicates the multiple regression model with the dependent variable PSAT Reading 2011 is a significant model ($p \leq .000$). The adjusted R^2 for this model indicates the model accounts for 18.7% of the variance in the dependent variable. The coefficients table (see Table 21) shows the predictor variable School was a significant contributor ($p \leq .002$). School had a positive β of .166 (2.8% of the variance) indicating that the school coded 1 (heterogeneous grouping school) had a weak positive effect on student performance on the reading portion of the 2011 PSAT. In addition to the variable of School, FRL (1.2% variance), Gender (1.4%), Absences (3.5%) and Grade 8 WASL reading (15.5%) were also significant contributors. Students who were not FRL (negative β), are males (negative β), had fewer absences (negative β) and scored higher on the Grade 8 WASL reading assessment (positive β) generally did better on the PSAT Reading 2011. Grade 8 WASL reading scores were the most significant predictor at 15.5% of the variance. This is 4.4 times stronger than the next strongest predictor (student absences).

PSAT Writing 2011

The PSAT in writing was administered to interested students on the same day as the PSAT in reading (October 2011). The ANOVA (see Table 24) indicates the multiple regression model with the dependent variable PSAT Writing is significant ($p \leq .001$). The adjusted R^2 for this model indicates the model accounts for 20.7% of the variance in the dependent variable.

The coefficients table (see Table 25) indicates that the predictor variable School was a significant contributor ($p \leq .003$). School had a positive β of .154 indicating that the school coded 1 had a weak positive effect on student performance on the writing portion of the 2011 PSAT (2.4% of variance). In addition to the variable of School, FRL (1.5% variance), Black/not Black (1.3%), Absences (1.9%), and Grade 8 WASL reading (17.7%) were significant contributors. Students who were not FRL (negative β), not black (negative β), had fewer absences (negative β) and scored higher on the Grade 8 WASL reading assessment (positive β) contributed positively to the dependent variable. Grade 8 WASL reading scores were the most significant predictor at 17.7% of the variance. This is 7.4 times stronger that the next strongest predictor (school).

This summary of Cohort 0 data allows an informed answer in regard to this study's research questions:

- To what extent, if any, does ability grouping of high achieving students defined as 9th and 10th grade honors English students in a suburban Washington State school district affect their performance on state and pre-college assessments of reading and writing achievement when controlling for student mutable variables?

Subsidiary Research Questions:

- To what extent, if any, does placement of 2009–2011 (Cohort 0) honors English students in homogeneous ability grouped English classes have on their subsequent performance on state and pre-college assessments of reading and writing achievement?
- To what extent, if any, does placement of 2009–2011 (Cohort 0) honors English students in heterogeneous ability groups have on their subsequent performance on state and pre-college assessments of reading and writing achievement?

Conclusion

The results of this study indicate that grouping Cohort 0 honors English students homogeneously according to ability does not significantly influence their performance in the models that were significant ($p \leq .05$), valid (adequate sample size), or where school was a significant contributor.

The results indicate that grouping Cohort 0 honors English students heterogeneously according to ability does significantly influence their performance on both models where the model and variable of school were significant and the sample size was valid.

Results: Cohort 1

Cohort 1 represents students who were in the 8th grade during the 2009–2010 school year. Each student completed the state assessment in reading during the spring of 2010. The students then experienced honors English instruction during the 2010–2011 and 2011–2012 school years—their 9th and 10th grade years. This honors English instruction was delivered either in homogeneous ability groups (School 0) or heterogeneous ability groups (School 1). Four hundred fifty-three of the students took the PSAT in the fall of their sophomore year (2011). Scores for the reading and writing portion of this PSAT assessment were collected. Many students completed the PSAT in the fall of their junior year (2012). Scores for the reading and writing portion of the PSAT assessment were collected. Students also took the High School Proficiency Exam (HSPE) in the spring of their sophomore year (2012). Student scores for the reading portion of this assessment were collected.

Tables 26–29 show the results of the regression analysis on the dependent/outcome variable HSPE Reading 2012 when the 2010 Grade 8 MSP Reading scores, Student absences,

Ethnicity (A-NA, B-NB, H-NH, OM-NOM), Special education, FRL, School, and Gender were used in the model as independent/predictor variables.

Table 26

Multiple regression variables entered HSPE reading 2012

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Grade_8_MSP_Reading_2010, Absences_by_Median, OM_NOM, Sped, H_NH, FRL, School, B_NB, Gender, A_NA[b]		Enter

a. Dependent Variable: Grade_10_HSPE_Reading_2012

b. All requested variables entered.

Table 27

Multiple regression model summary HSPE reading 2012

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.436[a]	.190	.171	23.056

a. Predictors: (Constant), Grade_8_MSP_Reading_2010, Absences_by_Median, OM_NOM, Sped, H_NH, FRL, School, B_NB, Gender, A_NA

The adjusted R^2 for this model indicates that 17.1% of the variance in student performance on the 2012 HSPE reading assessment can be explained by School (proxy for type of grouping), FRL, Gender, Ethnicity (Black/not Black, Asian/not Asian, Hispanic/not Hispanic, Other: multiracial/not other: multiracial), Special education, full day Absences, and Grade 8 WASL reading.

Table 28

Multiple regression ANOVA HSPE reading 2012

ANOVA[a]

Model		Sum of Squares	df	F	Sig.
1	Regression	52866.619	10	9.945	.000[b]
	Residual	225393.303	424		
	Total	278259.922	434		

a. Dependent Variable: Grade_10_HSPE_Reading_2012

b. Predictors: (Constant), Grade_8_MSP_Reading_2010, Absences_by_Median, OM_NOM, Sped, H_NH, FRL, School, B_NB, Gender, A_NA

The ANOVA table indicates the model is statistically significant $F = 9.945$, $df = 10, 424$, $p < .001$.

Table 29

Multiple regression coefficients table HSPE reading 2012

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Correlations			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	285.793	23.521		12.151	.000					
	School	8.741	2.364	.167	3.697	.000	.237	.177	.162	.940	1.063
	FRL	-6.113	2.764	-.102	-2.212	.028	-.165	-.107	-.097	.904	1.107
	Gender	.576	2.312	.011	.249	.803	.054	.012	.011	.925	1.081
	B_NB	-10.718	4.776	-.104	-2.244	.025	-.103	-.108	-.098	.890	1.124
	A_NA	-5.686	2.757	-.099	-2.062	.040	-.046	-.100	-.090	.827	1.210
	H_NH	-1.284	4.898	-.012	-.262	.793	-.019	-.013	-.011	.940	1.064
	OM_NOM	-.523	4.360	-.005	-.120	.905	.024	-.006	-.005	.943	1.060
	Sped	15.378	16.407	.041	.937	.349	.038	.045	.041	.992	1.008
	Absences_by_Median	-5.256	2.292	-.104	-2.293	.022	-.072	-.111	-.100	.931	1.074
	Grade_8_MSP_Reading_2010	.386	.056	.315	6.842	.000	.345	.315	.299	.900	1.111

a. Dependent variable: Grade_10_Reading_2012

Examination of the coefficients table indicates that there are six significant predictors: School $\beta = .167$, $t = 3.697$, $p < .001$: FRL $\beta = -.102$, $t = -2.212$, $p \leq .028$, Black/not Black $\beta = -.104$, $t = -2.244$, $p \leq .025$, Asian/not Asian $\beta = -.099$, $t = -2.062$, $p \leq .040$, Absences $\beta = -.104$, $t = -2.293$, $p \leq .022$; and Grade 8 MSP reading $\beta = .315$, $t = 6.842$, $p < .001$. School is a significant predictor with a positive β of .167. This indicates that School 1 (heterogeneous grouping) has a slightly positive effect on the dependent variable of HSPE Reading 2012. In

addition to school, the variable Grade 8 MSP reading (positive β) also positively contributes to the dependent variable. The negative β values of FRL, Black/not Black, Asian/not Asian and absences indicates that students who are not FRL, white, and those with fewer absences positively contribute to the model. Squaring the β values of each significant predictor indicates that School (proxy for the grouping variable) explains 2.8% of the model's variance, FRL 1.0%, Black/not Black 1.0%, Asian/not Asian 0.98%, Attendance 1.1%, and Grade 8 MSP reading 9.9%.

An analysis to determine the presence of multicollinearity issues is to examine both the tolerance and VIF values. According to Leech, Barrett, and Morgan (2011), multicollinearity may be an issue if tolerance values are lower than the value $(1 - R^2)$. The predictor Asian/not Asian has a tolerance value of .827, which is slightly lower than .829 (1 − .171). However, a further consideration is to examine the VIF. In this case, the VIF for A/NA is 1.210, which is close to one signifying a questionable although possible multicollinearity issue.

Tables 30–33 show the results of the regression analysis on the dependent/outcome variable PSAT Reading 2011 when the 2010 Grade 8 MSP Reading scores, Absences, Ethnicity (A-NA, B-NB, H-NH, OM-NOM), Special education, FRL, School, and Gender were used in the model as independent/predictor variables.

Table 30

Multiple regression variables entered PSAT reading 2011

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Grade_8_MSP_Reading_2010, Absences_by_Median, Sped, B_NB, H_NH, OM_NOM, Gender, School, FRL, A_NA[b]		Enter

a. Dependent variable: PSAT_Reading_2011
b. All requested variables entered

Table 31

Multiple regression model summary PSAT reading 2011

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.518[a]	.268	.250	7.047

a. Predictors: (Constant), Grade_8_MSP_Reading_2010, Absences_by_Median, Sped, B_NB, H_NH, OM_NOM, Gender, School, FRL, A_NA

The adjusted R^2 for this model indicates that 25.0% of the variance in student performance on the 2011 PSAT Reading assessment can be explained by School (proxy for type of grouping), FRL, Gender, Ethnicity (Black/not Black, Asian/not Asian, Hispanic/not Hispanic, Other: multiracial/not other: multiracial), Special education, Absences, and Grade 8 MSP Reading.

Table 32

Multiple regression ANOVA table PSAT reading 2011

ANOVA[a]

Model		Sum of Squares	df	F	Sig.
1	Regression	7448.725	10	14.999	.000[b]
	Residual	20311.808	409		
	Total	27760.533	419		

a. Dependent variable: PSAT_Reading_2011
b. Predictors: (Constant), Grade_8_MSP_Reading_2010, Absences_by_Median, Sped, B_NB, H_NH, OM_NOM, Gender, School, FRL, A_NA

The ANOVA table indicates the model is statistically significant $F = 14.999$, $df = 10, 409$, $p < .001$.

Table 33

Multiple regression coefficients table PSAT reading 2011

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-14.762	7.348		-2.009	.045					
	School	1.538	.736	.091	2.088	.037	.184	.103	.088	.942	1.062
	FRL	-3.415	.871	-.174	-3.921	.000	-.242	-.190	-.166	.905	1.106
	Gender	1.937	.716	.119	2.705	.007	.158	.133	.114	.931	1.074
	B_NB	-3.932	1.525	-.114	-2.578	.010	-.124	-.126	-.109	.908	1.101
	A_NA	-1.813	.854	-.099	-2.124	.034	-.037	-.104	-.090	.825	1.212
	H_NH	-.937	1.529	-.027	-.613	.540	-.032	-.030	-.026	.939	1.065
	OM_NOM	-.398	1.374	-.013	-.289	.772	.031	-.014	-.012	.945	1.059
	Sped	3.535	5.016	.030	.705	.481	.020	.035	.030	.992	1.008
	Absences_by_Median	-1.679	.711	-.103	-2.361	.019	-.066	-.116	-.100	.935	1.070
	Grade_8_MSP_Reading_2010	.151	.018	.382	8.550	.000	.432	.389	.362	.896	1.116

a. Dependent variable: PSAT_Reading_2011

Examination of the coefficients table indicates that there are seven significant predictors: School $\beta = .091$, $t = 2.088$, $p \leq .037$; FRL $\beta = -.174$, $t = -3.921$, $p < .001$; Gender $\beta = .119$, $t = 2.705$, $p \leq .007$; Black/not Black $\beta = -.114$, $t = -2.578$, $p \leq .010$; Asian/not Asian $\beta = -.099$, $t = -2.124$, $p \leq .034$; Absences $\beta = -.103$, $t = -2.361$, $p \leq .019$; and Grade 8 MSP reading $\beta = .382$, $t = 8.550$, $p < .001$. School is a significant predictor with a positive β of .091. This indicates that School 1 (heterogeneous grouping) has a very slight positive influence on the dependent variable

of PSAT Reading 2011. In addition to School, the variables with a positive β; Gender (female) and Grade 8 MSP reading (positive βs) positively contribute to the dependent variable. The negative β values of FRL, Black/not Black, Asian/not Asian, and attendance indicates that students that are not FRL, white, and those with fewer absences positively contribute to the model. Squaring the β values of each significant predictor signifies that School (proxy for the grouping variable) explains 0.8% of the model's variance, FRL 3.0%, Gender 4%, Black/not Black 1.3%, Asian/not Asian 0.98%, Attendance 1.1% and Grade 8 MSP reading 14.6%. An analysis of the model's tolerance and VIF values indicates no apparent issues with multicollinearity.

Tables 34–37 show the results of the regression analysis on the dependent/outcome variable PSAT Writing 2011 when the 2010 Grade 8 MSP Reading scores, student absences, Ethnicity (A-NA, B-NB, H-NH, OM-NOM), Special education, FRL, School, and Gender were used in the model as independent/predictor variables.

Table 34

Multiple regression variables entered/removed PSAT writing 2011

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Grade_8_MSP_Reading_2010, Absences_by_Median, Sped, B_NB, H_NH, OM_NOM, Gender, School, FRL, A_NA[b]		Enter

a. Dependent variable: PSAT_Writing_2011
b. All requested variables entered

Table 35

Multiple regression model summary PSAT writing 2011

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.513[a]	.263	.245	7.030

a. Predictors: (Constant), Grade_8_MSP_Reading_2010, Absences_by_Median, Sped, B_NB, H_NH, OM_NOM, Gender, School, FRL, A_NA

The adjusted R^2 for this model indicates that 24.5% of the variance in student performance on the 2011 PSAT Writing assessment can be explained by School (proxy for type of grouping), FRL, Gender, Ethnicity (Black/not Black, Asian/not Asian, Hispanic/not Hispanic, Other: multiracial/not other: multiracial), Special education, Absences, and Grade 8 MSP Reading.

Table 36

Multiple regression ANOVA table PSAT writing 2011

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7220.167	10	722.017	14.608	.000[b]
	Residual	20214.831	409	49.425		
	Total	27434.998	419			

a. Dependent Variable: PSAT_Writing_2011
b. Predictors: (Constant), Grade_8_MSP_Reading_2010, Absences_by_Median, Sped, B_NB, H_NH, OM_NOM, Gender, School, FRL, A_NA

The ANOVA table indicates the model is statistically significant $F = 14.608$, $df = 10, 409$, $p < .001$.

Table 37

Multiple regression coefficients table PSAT writing 2011

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-18.371	7.330		-2.506	013					
	School	1.615	.735	.096	2.198	.028	.180	108	.093	.942	1.062
	FRL	-3.614	.869	-.186	-4.160	.000	-.249	-.201	-.177	.905	1.106
	Gender	.909	.714	.056	1.273	.204	.097	.063	.054	.931	1.074
	B_NB	-3.498	1.521	-.102	-2.299	.022	-.126	-.113	-.098	.908	1.101
	A_NA	-.832	.852	-.046	-.977	.329	.016	-.048	-.041	.825	1.212
	H_NH	-.159	1.525	-.005	-.104	.917	-.019	-.005	-.004	.939	1.065
	OM_NOM	-.911	1.370	-.029	-.664	.507	.006	-.033	-.028	.945	1.059
	Sped	1.124	5.004	.010	.225	.822	-.003	.011	.010	.992	1.008
	Absences_by_Median	-2.190	.710	-.135	-3.086	002	-.115	-.151	-.131	.935	1.070
	Grade_8_MSP_Reading_2010	.151	.018	.386	8.619	.000	.431	.392	366	.896	1.116

a. Dependent Variable: PSAT_Writing_2011

Examination of the coefficients table indicates that there are five significant predictors: School $\beta = .096$, $t = 2.198$, $p \leq .028$; FRL $\beta = -.186$, $t = -4.160$, $p < .001$; Black/not Black $\beta = -.102$, $t = -2.290$, $p \leq .022$; Absences $\beta = -.135$, $t = -3.086$, $p \leq .002$; and Grade 8 MSP reading $\beta = .386$, $t = 8.619$, $p < .001$. School is a significant predictor with a positive β of .096. This indicates that School 1 (heterogeneous grouping) has a very slight positive influence on the dependent variable of PSAT Writing 2011. In addition to School, the variable Grade 8 MSP reading scores (positive β) positively contributes to the dependent variable. This means that higher scores on the MSP contribute to higher scores on the 2011 PSAT Writing section. The negative β values of FRL, Black/not Black, and Absences indicate that students who are not FRL, white, and those with fewer absences positively contribute to the model. Squaring the β values of each significant predictor signifies that school explains 0.92% of the model's variance, FRL 3.5%, Black/not Black 1.0%, Attendance 1.8%, and Grade 8 MSP reading 14.9%. An analysis of the model's tolerance and VIF values indicates no apparent issues with multicollinearity.

Tables 38–41 show the results of the regression analysis on the dependent/outcome variable PSAT Reading 2012 when the 2010 Grade 8 MSP Reading scores, Student attendance, Ethnicity (A-NA, B-NB, H-NH, OM-NOM), Special education, FRL, School, and Gender were used in the model as independent/predictor variables.

Table 38

Multiple regression variables entered PSAT reading 2012

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Grade_8_MSP_Reading_2010, B_NB, Absences_by_Median, OM_NOM, H_NH, School, FRL, Gender, A_NA[b]		Enter

a. Dependent Variable: PSAT_Reading_2012
b. All requested variables entered

Table 39

Multiple regression model summary PSAT reading 2012

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.530[a]	.280	.249	7.933

a. Predictors: (Constant), Grade_8_MSP_Reading_2010, B_NB, Absences_by_Median, OM_NOM, H_NH, School, FRL, Gender, A_NA

The adjusted R^2 for this model indicates that 24.9% of the variance in student performance on the 2012 PSAT Reading assessment can be explained by School (proxy for type of grouping), FRL, Gender, Ethnicity (Black/not Black, Asian/not Asian, Hispanic/not Hispanic, Other or multiracial/not other, or Multiracial), Special education, Absences, and Grade 8 MSP Reading.

Table 40

Multiple regression ANOVA PSAT reading 2012

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5027.089	9	558.565	8.877	.000[b]
	Residual	12899.683	205	62.925		
	Total	17926.772	214			

a. Dependent Variable: PSAT_Reading_2012
b. Predictors: (Constant), Grade_8_MSP_Reading_2012, B_NB, Absences_by_Median, OM_NOM, H_NH, School, Gender, FRL, A_NA

The ANOVA table indicates the model is statistically significant $F = 8.877$, $df = 9, 205$, $p < .001$.

Table 41

Multiple regression coefficients table PSAT reading 2012

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-4.533	11.732		-.386	.700					
	School	3.562	1.162	.188	3.066	.002	.265	.209	.182	.938	1.066
	FRL	-5.340	1.535	-.223	-3.479	.001	-.307	-.236	-.206	.854	1.171
	Gender	-.139	1.140	-.008	-.122	.903	.044	-.009	-.007	.911	1.097
	B_NB	-6.353	2.322	-.172	-2.736	.007	-.232	-.188	-.162	.892	1.122
	A_NA	.162	1.400	.008	.116	.908	.015	.008	.007	.794	1.259
	H_NH	1.296	2.563	.031	.506	.614	-.038	.035	.030	.918	1.090
	OM_NOM	.328	2.013	.010	.163	.871	.040	.011	.010	.942	1.062
	Absences_by_Med	-1.727	1.112	-.094	-1.552	.122	-.095	-.108	-.092	.961	1.040
	Grade_8_MSP_Reading_2010	.137	.028	.311	4.891	.000	.368	.323	.290	.869	1.151

a. Dependent variable: PSAT_Reading_2012

Examination of the coefficients table indicates that there are four significant predictors: School $\beta = .188$, $t = 3.066$, $p \leq .002$; FRL $\beta = -.223$, $t = -3.479$, $p \leq .001$; Black/not Black $\beta = -.172$, $t = -2.736$, $p \leq .007$, and Grade 8 MSP reading $\beta = .311$, $t = 4.891$, $p < .001$. School is a significant predictor with a positive β of .188. This indicates that School 1 (heterogeneous grouping) has a slight positive influence on the dependent variable of PSAT Reading 2012. In addition to school, the variable Grade 8 MSP reading scores (positive β) positively contributes to the dependent variable. This means that higher scores on the MSP contribute to higher scores on

the 2012 PSAT Reading section. The negative β values of FRL and Black/not Black indicate that students who are not FRL and white positively contribute to the model. Squaring the β values of each significant predictor signifies that school explains 3.5% of the model's variance, FRL 5.0%, Black/not Black 3.0%, and Grade 8 MSP scores 9.7%. An analysis of the model's tolerance and VIF values indicates no apparent issues with multicollinearity.

Tables 42–45 show the results of the regression analysis on the dependent/outcome variable PSAT Writing 2012 when the 2010 Grade 8 MSP Reading scores, student attendance, Ethnicity (A-NA, B-NB, H-NH, OM-NOM), Special education, FRL, School, and Gender were used in the model as independent/predictor variables.

Table 42

Multiple regression variables entered PSAT writing 2012

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Grade_8_MSP_Reading_2010, B_NB, Absences_by_Median, OM_NOM, H_NH, School, FRL, Gender, A_NA[b]		Enter

a. Dependent variable: PSAT_Writing_2012
b. All requested variables entered

Table 43

Multiple regression model summary PSAT writing 2012

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.502[a]	.252	.220	8.039

a. PredictorsL (Constant), Grade_8MSP_Reading_2012, B_NB, Absences_by_Median, OM_NOM, H_NH, School, FRL, Gender, A_NA

The adjusted R^2 for this model indicates that 22.0% of the variance in student performance on the 2012 PSAT Writing assessment can be explained by School (proxy for type of grouping), FRL, Gender, Ethnicity (Black/not Black, Asian/not Asian, Hispanic/not Hispanic, Other: multiracial/not other: multiracial), special education, number of absences, and Grade 8 MSP Reading.

Table 44

Multiple regression ANOVA PSAT writing 2012

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4472.165	9	496.907	7.688	.000[b]
	Residual	13249.667	205	64.633		
	Total	17721.833	214			

a. Dependent variable: PSAT_Writing_2012
b. Predictors: (Constant), Grade_8_MSP_Reading_2010, B_NB, Absences_by_Median, OM_NOM, H_NH, School, FRL, Gender, A_NA

The ANOVA table indicates the model is statistically significant $F = 7.688$, $df = 9, 205$, $p < .001$.

Table 45

Multiple regression coefficients table PSAT writing 2012

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-3.846	11.890		-.323	.747					
	School	3.330	1.177	.176	2.828	.005	.241	.194	.171	.938	1.066
	FRL	-5.381	1.556	-.226	-3.459	.001	-.279	-.235	-.209	.854	1.171
	Gender	1.195	1.155	.065	1.035	.302	.110	.072	.062	.911	1.097
	B_NB	-2.942	2.353	-.080	-1.250	.213	-.148	-.087	-.075	.892	1.122
	A_NA	.626	1.419	.030	.441	.659	.028	.031	.027	.794	1.259
	H_NH	2.112	2.598	.051	.813	.417	-.036	.057	.049	.918	1.090
	OM_NOM	1.431	2.040	.044	.702	.484	.056	.049	.042	.942	1.062
	Absences_by_Median	-2.457	1.127	-.134	-2.179	.030	-.129	-.150	-.132	.961	1.040
	Grade_8_MSP_Reading_2010	.129	.028	.294	4.542	.000	.370	.302	.274	.869	1.151

a. Dependent variable: PSAT_Writing_2012

Examination of the coefficients table indicates that there are four significant predictors: School $\beta = .176$, $t = 2.828$, $p \leq .005$; FRL $\beta = -.226$, $t = -3.459$, $p \leq .001$, Absences $\beta = -.134$, t

= −2.179, $p \leq .030$ and Grade 8 MSP reading $\beta = .294$, $t = 4.542$, $p < .001$. School is a significant predictor with a positive β of .176. This indicates that School 1 (heterogeneous grouping) has a positive influence on the dependent variable of PSAT Writing 2012. In addition to school, the variable Grade 8 MSP reading scores (positive β) also positively contributes to the dependent variable. This means that higher scores on the MSP contribute to higher scores on the 2012 PSAT Writing section. The negative β values of FRL and Attendance indicate that students who are not FRL, and those with fewer absences positively contribute to the model. Squaring the β values of each significant predictor signifies that School explains 3.1% of the model's variance, FRL 5.1%, Attendance 1.8%, and Grade 8 MSP reading 8.6%. An analysis of the model's tolerance and VIF values indicates no apparent issues with multicollinearity.

Summary of Results: Cohort 1

Students in Cohort 1 experienced honors English instruction for two consecutive years. Cohort 1 students were in the 8th grade during the 2009–2010 school year. Upon entering their first year of high school, the 9th grade, students chose honors English. In School 0, the curriculum was delivered in a homogeneous environment where all students in the honors English classroom were honors students. In School 1, the curriculum was delivered in a heterogeneous environment where approximately half the classroom students were honors students and the others were not ("core" English students). Students continued in these like-ability (School 0) or mixed-ability (School 1) honors English classes through their 10th grade year. Five state and/or pre-college assessments in reading and/or writing were offered.

High school proficiency exam (HSPE) in reading 2012

This state standardized assessment was taken by students on the same day in March, 2012; the spring of the students' sophomore year. The multiple regression ANOVA (see Table

28) for the model with the dependent variable HSPE Reading 2012 was significant ($p \leq .001$). The adjusted R^2 indicates this model accounts for 17.1% of the variance. The multiple regression coefficients (see Table 29) shows that School (proxy for grouping type) was a significant contributor to this model ($p \leq .000$). The variable School had a $\beta = .167$. This positive β indicates that School 1 (heterogeneously grouping school) had a weak but positive effect on a student's performance on the 2012 Reading HSPE (2.8% variance). The other significant predictors were Grade 8 MSP reading (positive β with 2.8% variance) and predictors with negative βs: FRL, Black/not Black, Asian/not Asian and number of Absences. Each of the significant predictors had a variance of 1.0%, 1.0 %, 0.98%, and 1.1% respectively. The variable Grade 8 MSP reading scores was the most significant predictor with 9.9% of the variance. This was 3.5 times stronger than the next highest predictor (School). Due to a low tolerance value for the predictor Asian/not Asian that is slightly lower than $(1 - R^2)$ a multicollinearity issue, although not likely, is possible.

PSAT Reading 2011

The PSAT in reading was administered to interested students in the fall of their sophomore year. The ANOVA (see Table 32) indicates that the multiple regression model with the dependent variable PSAT Reading 2011 was a significant model ($p < .001$). The adjusted R^2 indicates this model accounts for 25% of the variance. The multiple regression coefficients table (see Table 33) shows that School was a significant contributor ($p \leq .037$). The β for School was .091. This positive β indicates that School 1 (heterogeneous grouping school) had a very weak positive effect on a student's performance on the reading portion of the 2011 PSAT. The other significant predictors were FRL (negative β), Black/not Black (negative β), Asian/not Asian (negative β), Attendance (negative β), Gender (positive β), and Grade 8 MSP reading (positive

β). Therefore, students who were not free and reduced lunch, females (not males), students who were white, those with fewer Absences, and those with higher Grade 8 MSP scores positively contributed to the model. The variable Grade 8 MSP reading scores was the most significant predictor with 14.6% of the variance. This was nearly 4.9 times stronger than the next highest predictor (FRL). An analysis of the model's tolerance and VIF values indicates no apparent issues with multicollinearity.

PSAT Writing 2011

The PSAT in writing was administered to interested students in the fall of their sophomore year. The ANOVA (see Table 36) indicates that the multiple regression model with the dependent variable PSAT Writing 2011 was a significant model ($p \leq .001$). The adjusted R^2 indicates this model accounts for 24.5% of the variance. The coefficient table (see Table 37) shows the variable School was a significant predictor ($p \leq .028$). The β for school was .096. This positive β indicates that School 1 (heterogeneous grouping school) had a very weak positive effect on a student's performance on the writing portion of the 2011 PSAT. The other significant predictors were FRL (negative β), Black/not Black (negative β), Absences (negative β) and Grade 8 MSP reading (positive β). Therefore, students who were not free and reduced lunch, students who were white, those with fewer Absences, and those with higher-Grade 8 MSP scores positively contributed to the model. The variable Grade 8 MSP reading scores was the most significant predictor with 14.9% of the variance. This was nearly 4.3 times stronger than the next highest predictor (FRL). An analysis of the model's tolerance and VIF values indicates no apparent issues with multicollinearity.

PSAT Reading 2012

The PSAT in reading was administered to interested students in October 2012 of their junior year. The ANOVA table (see Table 40) indicates the multiple regression model with the dependent variable PSAT Reading 2012 is a significant model ($p < .000$). The adjusted R^2 indicates this model accounts for 24.9% of the variance. The coefficients table (see Table 41) shows the predictor variable School was a significant contributor ($p \leq .002$). School had a positive β of .188 indicating that the School 1 (heterogeneous grouping school) had a weak positive effect on student performance on the reading portion of the 2012 PSAT. The other significant predictors were FRL (negative β), Black/not Black (negative β), and Grade 8 MSP reading (positive β). Therefore, students who were not in the FRL program, students who were white and those with higher Grade 8 MSP scores positively contributed to the model. The variable Grade 8 MSP reading scores was the most significant predictor with 9.7% of the variance. This was nearly 1.9 times stronger than the next highest predictor (FRL). An analysis of the model's tolerance and VIF values indicates no apparent issues with multicollinearity exist.

PSAT Writing 2012

The PSAT in writing was administered to interested students on the same day as the PSAT in reading (October 2012). The ANOVA (see Table 44) indicates the multiple regression model with the dependent variable PSAT Writing is a significant model ($p < .001$). The adjusted R^2 indicates this model accounts for 22.0% of the variance. The coefficients table (see Table 45) indicates that the predictor variable School was a significant contributor ($p \leq .005$). School had a positive β of .176 indicating that the school coded 1 (heterogeneous grouping school) had a weak positive effect on student performance on the writing portion of the 2012 PSAT. The other significant predictors were FRL (negative β), Absences (negative β) and Grade 8 MSP reading

(positive β). Therefore, students who were not FRL, students who had fewer absences, and those with higher Grade 8 MSP scores positively contributed to the model. The variable Grade 8 MSP reading scores was the most significant predictor with 8.6% of the variance. This was nearly 1.7X stronger than the next highest predictor (FRL). An analysis of the model's tolerance and VIF values indicates no apparent issues with multicollinearity.

This summary of Cohort 1 data allows an informed answer in regard to this study's research questions.

- To what extent, if any, does ability grouping of high achieving students defined as 9th and 10th grade honors English students in a suburban Washington State school district affect their performance on state and pre-college assessments of reading and writing achievement when controlling for student mutable variables?

Subsidiary research questions:

- To what extent, if any, does placement of 2010–2012 (Cohort 1) honors English students in homogeneous ability grouped English classes have on their subsequent performance on state and pre-college assessments of reading and writing achievement?
- To what extent, if any, does placement of 2010–2012 (Cohort 1) honors English students in heterogeneous ability groups have on their subsequent performance on state and pre-college assessments of reading and writing achievement?

Conclusion

The results of this study indicate that grouping Cohort 1 honors English students homogeneously according to ability does not significantly influence their performance on any of the five reading or writing assessments used in this study.

The results of this study indicate that grouping Cohort 1 honors English students heterogeneously according to ability does significantly influence their performance on all five reading or writing assessments used in this study.

Summary of overall findings (across cohorts)

Correlation analysis.

The variable School (proxy for type of grouping) was not significantly correlated with any of the dependent variables in Cohort 0. School was weakly and significantly correlated with each of the dependent variables (ranging from Pearson's r = .176 to Pearson's r = .265) for Cohort 1. The weak correlations were all positive, indicating scores on the dependent variables are positively correlated with the school coded as 1 (heterogeneous grouping school).

Multiple regression models.

Model 1: High School Proficiency Exam (HSPE) in reading 2011 or 2012

In Cohort 0, School (proxy for grouping type) was not a significant contributor to the HSPE Reading 2011 model ($p \leq .062$) while it was significant for Cohort 1 ($p \leq .000$). The positive β indicates that School 1 (heterogeneous grouping school) had a weak yet positive effect on a student's performance on the 2012 Reading HSPE.

Model 2: PSAT Reading 2010 or 2011

For Cohort 0, the multiple regression model with the dependent variable PSAT Reading 2010 was not a significant model ($p \leq .068$). The regression for Cohort 1 (PSAT Reading 2011) had School (proxy for grouping type) as a significant contributor ($p \leq .037$). The positive β indicates that School 1 (heterogeneous grouping school) had a very weak yet positive effect on a student's performance on the reading portion of the 2011 reading portion of the PSAT.

Model 3: PSAT Writing 2010 or 2011

This particular model for Cohort 0 with the dependent variable of PSAT Writing 2010 had too few students (n = 38) take the assessment. Therefore, it does not warrant further consideration. The multiple regression for Cohort 1 with the dependent variable of PSAT Writing 2011 indicates School (proxy for grouping type) was a significant predictor ($p \leq .028$). The positive β indicates that School 1 (heterogeneous grouping school) had a very weak yet positive effect on a student's performance on the writing portion of the 2011 PSAT.

Model 4: PSAT Reading 2011 or 2012

This particular model for Cohort 0 shows the variable School (proxy for grouping type) was a significant contributor $p \leq .002$ and School had a positive β. This indicates that School 1 (heterogeneous grouping school) had a weak yet positive effect on student performance on the reading portion of the 2011 PSAT. This particular model for Cohort 1 shows the predictor variable School was a significant contributor ($p \leq .002$). School had a positive β indicating School 1 (heterogeneous grouping school) had a weak yet positive effect on student performance on the reading portion of the 2012 PSAT.

Model 5: PSAT Writing 2011 or 2012

This particular model for Cohort 0 indicates that the predictor variable School (proxy for grouping type) was a significant contributor ($p \leq .003$). School had a positive β indicating School 1 had a weak yet positive effect on student performance on the writing portion of the 2011 PSAT. This particular model for Cohort 1 indicates that the predictor variable School was a significant contributor ($p \leq .005$). School had a positive β indicating School 1 (heterogeneous grouping school) had a weak yet positive effect on student performance on the writing portion of the 2012 PSAT.

Conclusion

The results indicate that the grouping variable, represented by School, significantly contributes to seven of the ten multiple regression models. Furthermore, the results indicate that grouping honors English students homogeneously (School 0) does not significantly influence their performance on any of the multiple regression models. The results indicate that grouping honors English students heterogeneously (School 1) according to ability does significantly influence their performance on student achievement indicators at the secondary level, more so with Cohort 1 than Cohort 0. Also noteworthy is the finding that student performance in the 8^{th} grade assessment, which occurred prior to their placement in one of the two grouping arrangements, was the strongest predictor of performance. The evidence reported here suggests that the grouping variable (school) significantly contributed to seven of the ten regression models. Heterogeneous grouping significantly contributes while homogeneous grouping does not. See Table 46 and 47 for a comparison of the results reported in this chapter encompassing both cohort year groupings.

Table 46

Cohort 0: 2009–2011 multiple regressions summary

Model: Dependent Variable	Favored Grouping Type	Most Significant Predictor	No Significant Model
Model 1: HSPE 2011	Neither	Grade 8 WASL	
Model 2: PSAT (R) 2010			X
Model 3: PSAT (W) 2010			X
Model 4: PSAT (R) 2011	Heterogeneous (School 1) * Variance 2.8%	Grade 8 WASL	
Model 5: PSAT (W) 2011	Heterogeneous (School 1) * Variance 2.4%	Grade 8 WASL	

*denotes significant predictor variable at .05 (or .01) level of significance

Table 47

Cohort 1: 2010–2012 multiple regressions summary

Model: Dependent Variable	Favored Grouping Type	Most Significant Predictor	No Significant Model
Model 1: HSPE 2012	Heterogeneous (School 1) * Variance 2.8%	Grade 8 MSP	
Model 2: PSAT (R) 2011	Heterogeneous (School 1) * Variance 0.8%	Grade 8 MSP	
Model 3: PSAT (W) 2011	Heterogeneous (School 1) * Variance 0.9%	Grade 8 MSP	
Model 4: PSAT (R) 2012	Heterogeneous (School 1) * Variance 3.5%	Grade 8 MSP	
Model 5: PSAT (W) 2012	Heterogeneous (School 1) * Variance 3.1%	Grade 8 MSP	

*denotes significant predictor variable at .05 (or .01) level of significance

Chapter V

CONCLUSIONS AND RECOMMENDATIONS

Introduction

The 2001 No Child Left Behind Act (NCLB) substantially increased the testing requirements and set demanding accountability standards for schools, districts, and states (Bloomfield, 2003). To assess a school's progress in meeting standards, public school districts and their schools were asked to demonstrate that they were making Adequate Yearly Progress (AYP). AYP does not rely on a broad measure of a school's overall averages but requires schools to report on the achievement of a number of student subgroups to determine school effectiveness (Fusarelli, 2004). NCLB has as one of its goals to diminish the achievement gap between minority and nonminority children, as well as between low socio-economic students and their more affluent classmates (Bloomfield & Cooper, 2003; Day-Vines & Patton, 2003; Sunderman, 2003). By forcing schools and states to report out on the performance of each subgroup, NCLB has exposed the disparate academic achievement levels of our nation's students (Chambers, 2009; Giroux & Schmidt, 2004; Kress et al., 2011).

Public schools are examining their policies and instructional practices to address the achievement gap exposed by the reporting requirements of NCLB (Wenglinski, 2004). As accountability measures and stakes rise, there is a call for an improved use of scientific evidence to inform educational policymaking (Wiseman, 2010). In terms of the achievement gap, national studies at the secondary level show when students are grouped homogeneously according to ability there is a rise in achievement inequality between the groups (Gamoran & Mare, 1989; Hoffer, 1992). Grouping students homogeneously also results in having a disproportionate

number of minority and economically disadvantaged students in the lower ability groups (Ansalone 2006, 2009: Gamoran & Mare, 1989).

Ability grouping involves separating students into groups according to their perceived academic abilities (Biafora & Ansalone, 2008; Callahan, 2005; Cooper, 1999; Slavin, 1991). This separation can occur within classes or can be a structural adaptation in which students of higher academic ability are placed in classes separate from their lower performing peers (Ansalone & Biafora, 2010; Slavin, 1991). The students' prior academic achievement is usually the determining factor in whether students are placed in the higher functioning group or track (Archbald, Glutting, & Qian, 2009; Ballon, 2008; Slavin, 1991). These special classrooms for the higher performing (gifted or honors) students are accompanied by curricula and instructional practices different from the classes containing the lower functioning students. The objective of this difference is to provide a level of education commensurate with the high cognitive levels of gifted students (Ansalone & Biafora, 2010; Biafora & Ansalone, 2008; Preckel, Gotz, & Frenzel, 2010). Indeed, this grouping or tracking of the gifted students has empirical evidence of its benefits for the gifted student, and hence, is used to support its practice (Goldring, 1990; Kulik & Kulik, 1982; Rogers, 1993; Rubin, 2003; Shields, 2002). Advocates of this homogeneous ability grouping hold that teachers can best meet the needs of students whose abilities, motivation, and aspirations are similar (Allan, 1991; Benbow & Stanley, 1996; Gamoran, 2009; Oakes & Guiton, 1995).

On the other hand, this practice of homogeneous grouping has not generated nearly the amount of beneficial evidence for students in the lower ability groups. In fact, the practice of ability grouping has shown to depress the academic achievement of students placed in the lower groups (Ansalone, 2000; Carbonaro, 2005; Oakes, 1992; Oakes & Guiton, 1995; Slavin, 1991).

National studies at the secondary level reveal an increased gap in the achievement of students in the high and low ability groups (Callahan, 2005; Gamoran, Nystrand, Berends, & LePore, 1995; Hoffer, 1992).

When examining the students generally placed in each of these groups, one finds disproportionality. African-American, Hispanic, and students of low socio-economic status, for example, are more likely to be placed in lower ability groups (Ansalone 2001, 2003; Carbonaro & Gamoran, 2002; Goodlad & Oaks, 1988; Oakes, 1987). As many as 700 studies have explored the nature and consequences of tracking (Ansalone, 2006). Most show that tracking adversely affects the academic achievement and career paths of our disadvantaged students (Ansalone, 2006; Burris & Wellner, 2005; Gamoran & Mare, 1989; Hallinan & Kubitschek, 1999). In a study done by Chambers (2009), tracking or ability grouping practices stunted the achievement of students in the lower groups, thereby solidifying and intensifying disparities in performance between the groups. Therefore, the unequal allocation of instruction between these ability groups may result in the widening of the achievement gap between high and low level classes over time (Chambers, 2009; Gamoran & Mare, 1989; Gamoran et al., 1995). Despite the research, ability grouping remains a practice in about 80% to 85% of American high schools (Archbald et al., 2009). Therefore, if NCLB is asking educators to report on the progress of each subgroup of students, the practice of grouping our students by ability is problematic in getting all students to meet common standards.

The theoretical framework supporting this study suggests that grouping students according to ability facilitates the teaching process by making it easier for the teacher (Ansalone, 2009; Keliher, 1931). Other grouping arrangements may enhance learning but make teaching more difficult. Baines, Blatchford, and Kutnick (2003) summarized the dilemma; "achieving a

strategic balance is vital for effective teaching and learning but is one of the most difficult dilemmas facing teachers" (p. 10). Adding to the problem are findings that show segregating students according to achievement perpetuates the gap in academic achievement along race and class lines (Ansalone, 2000, 2004, 2009, 2010; Argys, Rees & Brewer, 1996; Burris & Wellner, 2005; Chambers, Higgins, & Scheurich, 2009; Gamoran & Mare, 1989; Gamoran et al., 1995; Gamoran & Carbonaro, 2003; Mallery & Mallery, 1999; Mickelson, 2001; Oakes, 1992).

The existing literature on the practice of tracking and/or ability grouping in our schools is ubiquitous. Despite the research, many parents and educators believe ability grouping benefits high achievers, and therefore, an entrenched culture of ability grouping remains firmly in place throughout America's high schools (Ansalone & Biafora, 2010; Biafora & Ansalone, 2008; Burris & Wellner, 2005; Preckel et al., 2010). These parents fear that if their honors students are placed with non-honors peers they will be exposed to an unchallenging curriculum, which will lower the standards for their children. Proponents of ability grouping, including the parents of honors students, want the honors students in ability groups (separate from the other students) so that they can reach their full potential without being hindered by the lower achieving students (Ansalone, 2010; Burris & Welner, 2005). Opponents of tracking, site research that shows the racial, ethnic, and SES disproportionality of students found in lower track (regular) courses vs. higher track (honors) courses. This disproportionality contributes to the achievement gap (Ansalone, 2006; Burris & Wellner, 2005; Livingston, 2010; Venzant, 2006). Furthermore, this arrangement hinders students who are in the lower track courses from meeting common standards and schools from meeting NCLB requirements.

Purpose

The purpose of this study was to determine if there is a tangible, measurable academic benefit to homogeneously grouping high school honors English students in a diverse, suburban school district in Washington State. This researcher examined specific models, which included the independent variables: state pre-assessment in reading, Gender, ELL, Ethnicity, Special education, student Attendance, and free and reduced lunch status. These variables combined with student placement in either homogeneous or heterogeneous ability groups were studied to determine what effect, if any, they had on the dependent variable of honors students' achievement as measured by state and pre-college assessments in reading and/or writing. The study is intended to produce evidence school district policymakers can use to make research-based decisions about ability grouping; decisions that will benefit all students.

At a time when the Common Core State Standards have been adopted in 45 states and all students leaving school college and career ready is an emphasis, the findings are a bit surprising. How can American schools accomplish the goal of getting *all* students to meet common standards? Some advocate homogeneous ability grouping so students can reach their maximum potential without being hindered by lower achieving students (Ansalone, 2009; Burris & Wellner, 2005). Others site the race, ethnic, and SES disproportionality of students found in lower ability groups and consider this a major contributor to these students not achieving at the same levels as their separately grouped peers (Ansalone, 2006; Burris & Wellner, 2005; Livingston, 2010; Venzant, 2006). Demonstrating an empirically verifiable benefit to the upper-end students is the major reason for homogeneously grouping students by ability since there is overwhelming evidence to support the notion that lower performing students' academic growth is hindered by their placement separate from higher performing honors students (Archbald &

Keleher, 2008; Hallinan & Kubitschek, 1999; Rubin, 2003; Rubin & Noguera, 2004). If there is no empirical evidence showing a benefit to homogeneous ability grouping of high school honors English students, why continue the practice?

The theoretical framework for this study supports the notion that instruction can be targeted more efficiently when students are grouped homogeneously (Allan, 1991; Barnard, as cited by Tyack 1974; Benbow & Stanley 1996; Gamoran 2009; Gamoran & Weinstein 1998: Keliher, 1931; Oakes & Guiton 1995; and Turney, 1931). Additionally, Hallinan and Sorensen (1983) referred to homogeneous grouping as the best way to manage students and keep them attentive; and Oakes (1987) stated "tracking...was adopted as the means for managing student diversity" (p. 129). Perhaps Slavin (1987) articulated the theoretical framework for homogeneous ability grouping best when he said it is supposed to:

> increase student achievement primarily by reducing the heterogeneity of the class or instructional group, making it more possible for the teacher to increase the pace and level of instruction for high achievers and provide more individual attention, repetition, and review for low achievers. (p. 296)

The research questions that follow were posed to determine the validity of claims such as these. Does homogeneity in instructional grouping measurably increase the pace and level of instruction for the high achievers as measured by subsequent achievement scores on state and pre-college measures of reading and/or writing?

Research question. To what extent, if any, does ability grouping of high achieving students defined as 9th and 10^{th} grade honors English students in a suburban Washington State school district affect their performance on state and pre-college assessments of reading and writing achievement when controlling for student mutable variables?

Subsidiary research questions

- To what extent, if any, does placement of 2009–2011 (Cohort 0) honors English students in homogeneous ability grouped English classes have on their subsequent performance on state and pre-college assessments of reading and writing achievement?
- To what extent, if any, does placement of 2009–2011 (Cohort 0) honors English students in heterogeneous ability grouped English classes have on their subsequent performance on state and pre-college assessments of reading and writing achievement?
- To what extent, if any, does placement of 2010–2012 (Cohort 1) honors English students in homogeneous ability grouped English classes have on their subsequent performance on state and pre-college assessments of reading and writing achievement?
- To what extent, if any, does placement of 2010–2012 (Cohort 1) honors English students in heterogeneous ability groups have on their subsequent performance on state and pre-college assessments of reading and writing achievement?

The results of this study do show that ability grouping high school honors English students in a suburban school district in Washington State affects their achievement on measures of state and pre-college reading and writing proficiency. The findings show that while grouping as a variable was a significant contributor, the type of grouping favored was heterogeneous grouping. This runs counter to much of the research reviewed in this study. It is often cited and believed that high achieving students benefit from being in classrooms with other high achievers, separated from their lower achieving peers. This study does not support this contention. Instead, it was the high achievers in heterogeneous classrooms that contributed positively to the outcomes reported on state and pre-college assessments. This was true across both cohorts. Each of the subsidiary research questions can now be answered.

Answers to subsidiary research questions. (a) Cohort 0 honors English high school students grouped homogeneously did not contribute positively to any of the dependent variables measuring student achievement in reading or writing achievement, (b) Cohort 0 honors English high school students grouped heterogeneously contributed positively to all significant models and to the models where the grouping variable was significant. The positive contribution was present for all studied dependent variables, (c) Cohort 1 honors English high school students grouped homogeneously did not contribute positively to any of the dependent variables measuring student achievement in reading and writing achievement, (d) Cohort 1 honors English high school students grouped heterogeneously contributed positively to all significant models and to the models where the grouping variable was significant. The positive contribution was present for all studied dependent variables.

Conclusions

This study was in effect two separate studies broken down into Cohort 0 and Cohort 1. Students in Cohort 0 were in the 8th grade during the 2008–2009 school year and were provided an honors English curriculum during their 9th and 10th grade years. Students in Cohort 1 were in the 8th grade during the 2009–2010 school year and were provided an honors English curriculum during their 9th and 10th grade years. Regardless of cohort, honors English students in School 0 were grouped homogeneously and honors English students in School 1 were grouped heterogeneously. Each cohort of students was assessed on five discrete dependent variables (state assessments in reading, and PSAT scores in reading and writing).

Results of this study suggest that the variable of ability grouping was a significant contributor in 7 out of 10 models (5 models per cohort). In terms of the three models where ability grouping was not significant, one model itself was not significant, one significant model

did not meet the sample size requirement advocated by Field (2009), and one significant model indicated that the ability grouping variable was not significant. In all seven significant models where ability grouping was significant, the type of grouping having a positive and significant influence on the respective dependent variable was heterogeneous ability grouping.

Of particular note, the percentage of variance in each of the models for students in Cohort 1 exceeded the variance in Cohort 0. This is perplexing. Perhaps the larger sample size of students in Cohort 1 (n = 474) vs. Cohort 0 (n = 388) added to the strength of the results and hence, the increase in variance. The fact is, the observable difference in percent variance across cohorts cannot be explained adequately.

The findings of this study challenge those of some previous empirical research. Argys et al. (1996) in a study of tracking's effect on 8^{th} and 10^{th} grade math students, found an 8.4% decline in student scores when high achieving students were moved from homogeneous ability classrooms to heterogeneous ones. Similarly, Goldring (1990) found gifted students in homogeneous math and science classes outperformed their gifted peers in heterogeneous classes. Goldring (1990) found a smaller benefit in reading and writing classes. Likewise, Allan (1991) and Shields (2002) found higher performing students benefited from being segregated from their lower performing peers.

In this study, it was discovered that although grouping was a significant variable in the regression models, it was heterogeneous ability grouping that contributed positively. In fact, in no model did homogeneous grouping positively contribute. This raises an interesting question. Why did honors students do better when grouped with non-honors students? Vygotsky (1978) reasoned that learning cannot be separated from a social context. Perhaps, as Vygotsky (1978) recognized, the interactions occurring between the honors and non-honors students contributed to

greater sense making and knowledge construction. These interactions may have allowed for a richer, and hence, more valuable learning experience for both high and lower achievers. This reciprocity or back and forth communication between students enables them to act as teachers of one another. It is no secret that to teach another is to become a better learner oneself.

Also worth considering is the fact that a teacher, in order to meet the needs of the wide range of learners, needs to employ a variety of instructional strategies. Doing so requires more effort by the teacher, which may cause him or her to be more consistently active in the instructional process. This increased teacher instructional activity may then heighten the engagement of the students in the classroom. Additionally, the teachers involved with teaching heterogeneously grouped students may have been exposed to an increased level of professional development. Knowing that students with varying levels of prior achievement need instruction, which is differentiated, the teachers may have received an increased amount of training on how to differentiate their instruction. This differentiation may have manifested itself as increased scaffolding of instruction. This means the teacher may have made intentional efforts to support the student at his or her current learning level and then provided experiences that move the student from one learning level to increasingly higher levels of cognition.

Because homogeneous grouping of high performing students is still a widespread practice in American high schools, the results of this study are noteworthy. The fact that this study's high school honors English students in heterogeneous classrooms did as well or better on measures of reading and writing proficiency compared with their homogeneously grouped peers warrants attention. The practice of grouping high school honors English students homogeneously may not be more advantageous to their academic success than placing them in heterogeneous classrooms. This study raises the question that type of ability grouping may be less important than what is

occurring in the classroom. Research is clear on the negative effects of homogeneous grouping on the lower performing students (Carbonaro, 2005; Hoffer & Gamoran, 1993; Hooper & Hannafin, 1991). This study raises an important issue. Is the type of grouping secondary to other classroom characteristics? Quality student learning experiences are possible in a grouping arrangement (heterogeneous) that does not leave the non-honors (lower performing) student behind. Tables 48 and 49 replicate the summary information provided in chapter four of this study.

Table 48

Cohort 0: 2009–2011 multiple regressions summary

Model: Dependent Variable	Favored Grouping Type	Most Significant Predictor	No Significant Model
Model 1: HSPE 2011	Neither	Grade 8 WASL	
Model 2: PSAT (R) 2010			X
Model 3: PSAT (W) 2010			X
Model 4: PSAT (R) 2011	Heterogeneous (School 1) * Variance 2.8%	Grade 8 WASL	
Model 5: PSAT (W) 2011	Heterogeneous (School 1) * Variance 2.4%	Grade 8 WASL	

*denotes significant predictor variable at .05 (or .01) level of significance

Table 49

Cohort 1: 2010–2012 multiple regressions summary

Model: Dependent Variable	Favored Grouping Type	Most Significant Predictor	No Significant Model
Model 1: HSPE 2012	Heterogeneous (School 1) * Variance 2.8%	Grade 8 MSP	
Model 2: PSAT (R) 2011	Heterogeneous (School 1) * Variance 0.8%	Grade 8 MSP	
Model 3: PSAT (W) 2011	Heterogeneous (School 1) * Variance 0.9%	Grade 8 MSP	
Model 4: PSAT (R) 2012	Heterogeneous (School 1) * Variance 3.5%	Grade 8 MSP	
Model 5: PSAT (W) 2012	Heterogeneous (School 1) * Variance 3.1%	Grade 8 MSP	

*denotes significant predictor variable at .05 (or .01) level of significance

The results of this study suggest homogeneously grouping students by ability is not necessary to achieve the goal of challenging and better preparing the honors students for success on measures of reading and writing achievement. In fact, homogeneous grouping of honors English students did not significantly contribute to any of the significant models used in this study. Results from this study imply students in heterogeneous ability groups can be taught in a way that allows them to perform on measures of state and pre-college reading achievement at levels equal to or better than their homogeneously grouped peers.

The common perception, however, is that having mixed ability groups in classrooms lowers the expectations and standards for the high achieving students. Parents, most notably those whose children are in the upper track, were almost unanimous in their support of tracking as an instructional practice (Ansalone, 2010). Many parents assume that homogeneous grouping benefits high achievers. This is partly due to parents' perceptions that detracking (mixed ability

groups) will dilute the curriculum and lower the learning standards for their children (Burris & Wellner, 2005; Keller, 2011; Oakes & Guiton, 1995). Teachers rely on the theory of instructional efficiency because they have found that creating lessons to meet the needs of learners with varying needs is difficult. Argys et al. (1996) indicated teachers see tracking "as a way to reduce the range of performance and motivation...making teaching easier and preventing less able students from 'holding back' those with greater academic talent" (pp. 624-625). The results from this study suggest that educational leaders should reexamine homogeneous ability grouping and look for structures that will support high levels of learning for all students.

Recommendations for policy and practice

Research is clear that separating students into like-ability groups produces inequalities in student educational outcomes between groups (Gamoran, 1993; Hoffer & Gamoran, 1993; Mickelson, 2001). These inequalities create differences across groups in the quality and quantity of instruction (Gamoran, 1989; Oakes, 1985). Due to pressure from parents and teachers, however, school leaders and policymakers continue to separate highly capable students into their own classrooms or programs. The results of this study do not support the perception that high achieving students benefit academically when separated from their lower achieving peers. Educational leaders must challenge perceptions of ability grouping advocates, which are based on years of practice and experience. Our school leaders must have the courage to step up and advocate for the learning needs of all students. Teachers can be taught to differentiate their instruction effectively to meet the needs of the high achiever. Although this presents a greater challenge for the teacher, a teacher who is trained in the art of differentiating instruction can meet the challenge. Hattie (2002) completed a study where he arrives at the powerful conclusion that "whether a school tracks by ability or not...appears less consequential than whether it

attends to the nature and quality of instruction in the classroom." "The learning environments within the classroom, and the mechanisms and processes of learning that they foster, are by far the more powerful" (p. 449).

Additionally, the achievement gap is clearly defined and it is persistent. The fact that students of color and low socio-economic status are more often placed into the lower student groups, impinges on their achievement and their academic and social opportunities (Ansalone, 2000, 2004, 2006; Burris & Wellner, 2005; Gamoran, 1989, 1993; Gamoran, et al., 1995; Oakes & Wells, 1998). The practice of tracking then widens the gap in achievement between the high and low track students, and therefore, between students who are poor, black, or Latino and those who are wealthier, white, or Asian (Ballon, 2008; Gamoran & Carbonaro, 2003; Mallory & Mallory, 1999; Mickelson, 2001: Oakes & Lipton, 1992; Oakes, 1992).

Educational leaders would be wise to reexamine the notion that the homogeneous grouping of students by ability is sound practice. NCLB and the adoption of the Common Core State Standards reference the fact that we want all of our students to meet rigorous standards and be provided with instruction that will facilitate their meeting of these standards. On measures of state and pre-college assessments of reading and writing achievement, this study indicates that honors English students in heterogeneous ability groups in a suburban school district in Washington State can achieve at levels equal to or greater than students who are homogeneously grouped. This is an important and significant finding. School leaders need to confront the notion that separating students by ability in high school English classes is fair and equitable. Prior research is clear that students in the lower ability classes do not benefit socially or academically. In fact, the achievement gap between them and their higher achieving peers grows. The results of this study suggest that the perceptions of homogeneous grouping

advocates are unwarranted. High achieving high school English students grouped heterogeneously can perform equally well or perhaps better than their homogeneously grouped peers on state and pre-college measures of reading and writing achievement. School leaders should reexamine the practice of grouping students homogeneously as a means for meeting the needs of the higher achievers. Instead, examining the instructional practices occurring in heterogeneous and homogeneous classrooms may uncover pedagogical strategies that are effective in meeting the learning needs of all students while grouped together. The answer to challenging the high achiever may not lie in the type of grouping but in the instructional practices occurring within the group.

Further recommendations for policy and practice

The findings from this study should be shared with teachers, parents, administrators, and policymakers. In one suburban school district in Washington State, high achieving high school English students grouped heterogeneously did as well as, or possibly better than, their peers grouped homogeneously as measured by state and pre-college assessments in reading and/or writing. The perception of homogeneous ability grouping advocates must be challenged. It is not necessary to group honors English students homogeneously in order to provide students with educational experiences, which enable them to achieve at high levels.

The results from this study add to the existing body of literature regarding the influence of grouping students according to ability and its effect on the academic achievement of honors English students. This study provides clear results. In 7 of the 10 multiple regression models, the grouping variable (represented by school) was a significant predictor. The students taught honors English in heterogeneous groups positively contributed to their reading and writing achievement as measured by each of the dependent variables. In one model, the type of grouping

was not significant and the other two were discarded due to lack of significance ($p > .05$) or an inadequate number of students taking the targeted assessment. This study provides additional insight in whether students should be placed in homogeneous groupings. It suggests that in order for schools and districts to meet the mandate set forth by NCLB and the Common Core State Standards placing students into heterogeneous groups has the potential to be just as effective for high achieving (honors) student as it is for the lower achieving student.

Recommendations for future research

This study provided empirical evidence about the effects of ability grouping high school honors English students in a suburban school district in Washington State. The results add to the existing body of research surrounding ability grouping (homogeneous vs. heterogeneous) and its influence on measures of reading and/or writing achievement. The small sample size and the purposeful makeup of the sample used in this study, does not allow generalizability to the wider population of schools or students. However, it does lead one to question whether the perceptions surrounding ability grouping and its effects on the high achiever are sound and research-based. It is important, therefore, to conduct future research in the area of ability grouping of high school students in order to determine if grouping students heterogeneously can positively affect the learning and achievement of all students. The following are this researcher's recommendations for further study:

- Replicate this study with a larger sample of students in another urban, suburban, or rural school district.
- Conduct a study where not only are students the unit of measurement but one that includes specific school variables that can be attributed to student academic success (e.g., teacher years of experience and highest degree attained, class size).

- Conduct a study of the instructional practices occurring within the classrooms of both homogeneous and heterogeneous ability classrooms in order to ascertain whether it's the independent variable of ability grouping or the strategies used in instruction that positively contributes the most to measures of academic achievement.
- Conduct a longitudinal study of longer than two years to determine if the findings from this study can be replicated or even accentuated by examining a longer period.
- Design a study to examine honors English teachers' attitudes toward the heterogeneous grouping of honors students and what support they would need to successfully teach in such an environment.
- Conduct a similar study, which also examines the performance of the non-honors students in the different grouping arrangements.
- Replicate the study in subjects other than English.

The requirement that all students regardless of Gender, Ethnicity, SES, ELL or Special education status meet the same rigorous standards appears to be a reality for years to come. A wealth of research supports the notion that students in the lower ability classrooms do not receive an education commensurate with their peers separated into higher achieving ones. Knowing this, further research into the practice of ability grouping and its effects on the high achiever is necessary and critical. If further research confirms the findings of this study, why continue the practice of homogeneous ability grouping?

REFERENCES

Agee, J. (2000). What is effective literature instruction? A study of experienced high school English teachers in differing grade and ability-level classes. *Journal of Literacy Research, 32*, 303-348. Retrieved from http://dx.doi.org/10.1080/10862960009548084

Allan, S. D. (1991). Ability-grouping research reviews: What do they say about grouping and the gifted? *Educational Leadership, 48*, 60-65.

Alvarez, D. & Mehan, H. (2005). Whole-school detracking: A strategy for equity and excellence. *Theory Into Practice, 45*, 82-89.

Ansalone, G. (2000). Keeping on track: A reassessment of tracking in the schools. *Race, Gender & Class, 7*, 108.

Ansalone, G. (2003). Poverty, tracking, and the social construction of failure: International perspectives on tracking. *Journal of Children & Poverty, 9*, 3-20.

Ansalone, G. (2004). Educational opportunity and access to knowledge: Tracking in the US and Japan. *Race, Gender & Class, 11*, 140.

Ansalone, G. (2006). Tracking: A return to Jim Crow. *Race, Gender & Class, 13*, 144-153.

Ansalone, G. (2009). Tracking, schooling and the equality of educational opportunity. *Race, Gender & Class, 16*, 174-185.

Ansalone, G. & Biafora, F. A. (2004). Elementary school teachers' perceptions and attitudes to the educational structure of tracking. *Education, 125*, 249-258.

Ansalone, G. & Biafora, F. A. (2010). Tracking in the schools: Perceptions and attitudes of parents. *Race, Gender & Class, 17*, 226-239.

Ansalone, G. & Ming, C. (2006). Programming students for academic success: The PLS an alternative to traditional tracking. *Educational Research Quarterly, 29*.

Archbald, D., Glutting, J., & Qian, X. (2009). Getting into honors or not: An analysis of the relative influence of grades, test scores, and race on track placement in a comprehensive high school. *American Secondary Education, 37*, 65-81.

Archbald, D. & Keleher, J. (2008). Measuring conditions and consequences of tracking in the high school curriculum. *American Secondary Education, 36*, 26-42.

Argys, L. M., Rees, D. I., & Brewer, D. J. (1996). Detracking America's schools: Equity at zero cost? *Journal of Policy Analysis and Management, 15*, 623-645.

Attewell, P. & Domina, T. (2008). Raising the bar: Curricular intensity and academic performance. *Educational Evaluation and Policy Analysis, 30*, 51-71.

Ayres, L. P. (1909). *Laggards in our schools: A study of retardation and elimination in city school systems*. New York, NY: Charities Publications Committee.

Baguley, T. (2004). Understanding statistical power in the context of applied research. *Applied Ergonomics, 35*, 73-80. doi:10.1016/j.apergo.2004.01.002

Baines, E., Blatchford, P., & Kutnick, P. (2003). Changes in grouping practices over primary and secondary school. *International Journal of Educational Research, 39*, 9-34.

Ballon, E. G. (2008). Racial differences in high school math track assignment. *Journal of Latinos and Education, 7*, 272-287. doi: 1080/15348430802143428

Barkley, R. (1955). The theory of the elite and the mythology of power. *Science & Society, 19*, 97-106.

Benbow, C. P., & Stanley, J. C. (1996). Inequity in equity: How "equity" can lead to inequity for high-potential students. *Psychology, Public Policy, and Law, 2*, 249-292.

Biafora, F. & Ansalone, G. (2008). Perceptions and attitudes of school principals towards school tracking: Structural considerations of personal beliefs. *Education, 128*(4), 588-602.

Bloomfield, D. C. & Cooper, B. S. (2003). NCLB: A new role for the federal government an overview of the most sweeping federal education law since 1965. *THE Journal, 30.*

Boaler, J. (2007). How a detracked mathematics approach promoted respect, responsibility, and high achievement. *Theory Into Practice, 45*, 40-46.

Bruner, J. S. (1963). *The process of education* (1st Vintage ed.). New York, NY: Random House.

Burr, M. Y. (1931). *A study of homogeneous grouping: In terms of individual variations and the teaching problem*. New York, NY: Teachers College, Columbia University.

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

Bybee, R. W. & Stage, E. (2005, 01/01/2005). No country left behind: International comparisons of student achievement tell U.S. educators where they must focus their efforts to create the schools the country needs. *Issues in Science and Technology, 21*. Retrieved from

CRA, P.L. § 88-352 (1964).

Caldas, S. J. (1993). Reexamination of input and process factor effects on public school achievement. *The Journal of Educational Research, 86*(4), 206-214.

Callahan, R. M. (2005). Tracking and high school English learners: Limiting opportunity to learn. *American Educational Research Journal, 42*, 305-328.

Campbell, D. T. & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Hopewell, N.J.: Houghton Mifflin.

Carbonaro, W. (2005). Tracking, students', effort, and academic achievement. *Sociology of Education, 78*, 27-49. doi: 10.1177/003804070507800102

Carbonaro, W. J. & Gamoran, A. (2002). The production of achievement inequality in high school English. *American Educational Research Journal, 39*, 801-827.

Cavanaugh, S. (2012). Complex policy options abound amid international comparisons. *Education Week, 31.*

Chambers, T. V. (2009, Fall). The "receivement gap": School tracking policies and the fallacy of the "achievement gap". *The Journal of Negro Education, 78,* 417-431.

Chambers, T. V., Higgins, K. S., & Scheurich, J. J. (2009). To track or not to track: Curricular differentiation and African American students at Highview High School. *Journal of Cases in Educational Leadership, 12,* 38-50. Retrieved from http://dx.doi.org/10.1177/1555458909334102

Chayte, M. (2009-2010). Thirty-five years after Berkelman: Seeking a new debate about ability grouping. *Hastings Constitutional Law Quarterly, 37,* 617-639. Retrieved from http://heinonline.org

Cheng, R. W., Lam, S.-F., & Chan, J. C. (2008). When high achievers and low achievers work in the same group: The roles of group heterogeneity and processes in project-based learning. *British Journal of Educational Psychology, 78,* 205-221.

Cohen, E. G. & Lotan, R. A. (1995). Producing equal-status interaction in the heterogeneous classroom. *American Educational Research Journal, 32,* 99-120.

Common Core State Standards Initiative website. (n.d.). Retrieved from http://www.corestandards.org/

Conant, J. B. (1959). *The American high school today* (1st ed.). New York, NY: McGraw Hill Book Company.

Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago, IL: Rand McNally.

Cooper, R. (1996). Detracking reform in an urban California high school: Improving the schooling experiences of African American students. *The Journal of Negro Education, 65*, 190-208.

Cooper, R. (1999). Urban school reform: Student responses to detracking in a racially mixed high school. *Journal of Education for Students Placed at Risk, 4*, 259-275.

Day-Vines, N. L., & Patton, J. M. (2003). The perils, pitfalls, and promises of the No Child Left Behind Act of 2001. *T/TAC Link Lines*, 1-5.

DiMartino, J. & Miles, S. (2005). Reaching real equity in schools. *Principal Leadership.*

Education for All Handicapped Children Act, § 94-192 (1975).

Educational Testing Service. (2010). *Washington assessment of student learning technical report 2009*. Retrieved from http://www.k12.wa.us/assessment/pubdocs/2009G8WASLTechReport.pdf

Educational Testing Service. (2011). *Washington assessment program grades 3-8, high school technical report*. Retrieved from http://www.k12.wa.us/assessment/TechReports.aspx

Educational Testing Service. (2012). *Washington assessment program grades 3-8, high school technical report*. Retrieved from http://www.k12.wa.us/assessment/TechReports.aspx

Epple, D., Newlon, E., & Romano, R. (2002). Ability tracking, school competition and the distribution of educational benefits. *Journal of Public Economics, 83*, 1-48.

Feldhusen, J. F. & Moon, S. M. (1992). Grouping gifted students: Issues and concerns. *Gifted Child Quarterly, 36*, 63-66.

Feldt, L. S. & Brennan, R. L. (1989). *Educational Measurement* (3rd ed.). New York, NY: MacMillan.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Findley, W. (1989). Tracking inequality: The problem with public school tracking. *Harvard Law Review, 102*, 1318-1341.

Freedman, S. W., Delp, V., & Crawford, S. M. (2005). Teaching English in untracked classrooms. *Research in the Teaching of English, 40*(1), 62-65.

Friedman, T. L. (2005). *The world is flat: A brief history of the 21st century* (1st ed.). New York, NY: Farrar, Straus and Giroux.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Karns, K. (1998). High achieving students' interactions and performance on complex mathematical tasks as a function of homogeneous and heterogeneous pairings. *American Educational Research Journal, 35*, 227-267.

Fusarelli, L. D. (2004). The potential impact of the No Child Left Behind Act on equity and diversity in American education. *Educational Policy, 18*, 71-94. doi: 10.1177/0895904803260025

Gallagher, J. (1997). Challenge or boredom? Gifted students' views on their schooling. *Roeper Review, 19*.

Gallaher, C. M. (2009). *A tale of two tracks: Analysis of student dialogue in standard and honors high school English classes* (Doctoral dissertation). Available from ProQuest. (UMI Number: 3356160)

Gamoran, A. (1989). *Instructional organization and educational equity*. National Center on Effective Secondary Schools, Madison, WI: Office of Educational Research and Improvement.

Gamoran, A. (1993). Alternative uses of ability grouping in secondary schools: Can we bring high-quality instruction to low ability classes? *American Journal of Education, 102.*

Gamoran, A. (2009). *Tracking and inequality: New directions for research and practice* (WCER Working Paper No. 2009-6) Madison: University of Wisconsin-Madison, Wisconsin Center for Education Research. Retrieved July 21, 2011, from http://www.wcer.wisc.edu/publications/workingpapers/papers.php

Gamoran, A. & Berends, M. (1987). The effects of stratification in secondary schools: Synthesis of survey and ethnographic research. *Review of Educational Research, 57,* 415-435. Retrieved from http://dx.doi.org/10.3102/00346543057004415

Gamoran, A. & Carbonaro, W. J. (2003). High school English: A national portrait. *High School Journal, 86.*

Gamoran, A. & Mare, R. D. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality? *American Journal of Sociology, 94,* 1146-1183. Retrieved from http://www.jstor.org/stable/2780469

Gamoran, A., Nystrand, M., Berends, M., & LePore, P. C. (1995). An organizational analysis of the effects of ability grouping. *American Educational Research Journal, 32,* 687-715.

Gamoran, A., & Weinstein, M. (1998). Differentiation and opportunity in restructured schools. *American Journal of Education, 106,* 385-415.

Gay, L. R., Mills, G. E., & Airasian, P. (2009). *Educational research: Competencies for analysis and applications* (9th ed.). Upper Saddle River, NJ: Merill.

George, D. & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference.* (4th ed.). Boston, MA: Allyn & Bacon.

Gerring, J. (2011). How good is good enough? A multidimensional, best-possible standard for research design. *Political Research Quarterly*, *64*, 625-636.

Giroux, M. & Schmidt, H. A. (2004). Closing the achievement gap a metaphor for children left behind. *Journal of Educational Change*, *5*, 213-228. doi: 10.1023/B:JEDU.0000041041.71525.67

Goldring, E. B. (1990). Assessing the status of information on classroom organizational frameworks for gifted students. *Journal of Educational Research*, *83*, 313-326.

Goodlad, J. I. & Oakes, J. (1988). We must offer equal access to knowledge. *Educational Leadership*, *45*, 16-22. Retrieved from http://www.eric.ed.gov/

Grunwald, H. E. & Mayhew, M. J. (2008). Using propensity scores for estimating causal effects:A study in the development of moral reasoning. *Res High Edu*, *49*, 758-775. doi:10.1007/s11162-008-9103-x

Haberman, M. (1991). The pedagogy of poverty versus good teaching. *Phi Delta Kappan*, *73*, 290-294.

Hallam, S. & Ireson, J. (2007). Secondary school pupils' satisfaction with their ability grouping placements. *British Educational Research Journal*, *33*, 27-45.

Hallinan, M. T. & Kubitschek, W. N. (1999). Curriculum differentiation and high school achievement. *Social psychology of education*, *3*, 41-62.

Hallinan, M. T. & Sorensen, A. B. (1983). The formation and stability of instructional groups. *American Sociological Association*, *48*, 838-851. Retrieved from http://www.jstor.org/stable/2095329

Hattie, J. A. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, *37*, 449-481. doi: 10.1016/S0883-0355(03)00015-6

Hewitt, T. W. (2008). Speculations on A Nation at Risk: Illusions and realities. *PHI DELTA KAPPAN*, 575-579.

Hoffer, T. B. (1992). Middle school ability grouping and student achievement in science and mathematics. *Educational Evaluation and Policy Analysis, 14*, 205-227. http://dx.doi.org/10.3102/01623737014003205

Hoffer, T. B. & Gamoran, A. (1993). *Effects of instructional differences among ability groups on student achievement in middle school science and mathematics*. Center on Organization and Restructuring of Schools, Northern Illinois University, Madison, WI: National Science Foundation.

Hooper, S. & Hannafin, M. (1991). The effects of group composition on achievement, interaction, and learning efficiency during computer-based cooperative instruction. *Educational Technology Research and Development, 39*(3), 27-40.

Ireson, J. & Hallam, S. (2005). Pupils' liking for school: Ability grouping, self-concept and perceptions of teaching. *British Journal of Educational Psychology, 75*, 297-311.

Ireson, J., Hallam, S., Hack, S., Clark, H., & Plewis, I. (2002). Ability grouping in English secondary school: Effects on attainment in English, math and science. *Educational Research and Evaluation: An Educational Journal on Theory and Practice, 8*, 299-318.

Ireson, J., Hallam, S., & Hurley, C. (2005). What are the effects of ability grouping on GCSE attainment? *British Educational Research Journal, 31*, 443-458.

Ireson, J., Hallam, S., Mortimer, P., Hack, S., & Clark, H. (1999). *Ability grouping in schools: An analysis of effects*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Jimenez-Buedo, M. & Miller, L. M. (2010). Why a trade-off? The relationship between the external and internal validity of experiments. *Theoria, 69*, 301-321.

Keliher, A. V. (1931). *A critical study of homogeneous grouping*. New York, NY: Teachers College, Columbia University.

Keller, T. T. (2011). *Ability grouping: The perceptions of the administrators, teachers, and parents* (Doctoral dissertation). Available from ProQuest. (UMI Number: 3445239)

Kerckhoff, A. C. (1986). Effects of ability grouping in British secondary schools. *American Sociological Review, 51*, 842-858.

Khazaeenezhad, B., Barati, H. & Jafarzade, M. (2012). Ability grouping as a way toward more academic success in teaching EFL—A case of Iranian undergraduates. *English Language Teaching, 5*, 81-89.

Kober, N., McMurrer, J., & Silva, M. R. (2011). *State test score trends through 2008-09, part 4: Is achievement improving and are gaps narrowing for Title I students?* Retrieved from Center on Education Policy: http://www.cep-dc.org

Kress, S., Zechmann, S., & Schmitten, J. M. (2011). Symposium: Education reform: When performance matters: The past, present, and future of consequential accountability in public education. *Harvard Journal on Legislation, 48*, 185.

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

LaPrade, K. (2011). Removing instructional barriers: One track at a time. *Education, 131*, 740-752.

LaPrade, K. A. (2008). *What happens when our students miss the "A" train: A detracking study to decrease the persistent achievement gap* (Doctoral dissertation). Available from ProQuest. (UMI Number: 3297703)

Leech, N. L., Barrett, K. C., & Morgan, G. A. (2011). *IBM SPSS for intermediate statistics* (4th ed.). New York, NY: Routledge.

Lemann, N. (1999). *The big test: The secret history of American meritocracy* (1st ed.). New York, NY: Farrar, Straus and Giroux.

Livingston, C. C. (2010). *The impact of tracking on student achievement: A multilevel growth model* (Doctoral dissertation). Available from ProQuest. (UMI Number: 3404161)

Lou, Y., Abrami, P. C., Spense, J. C., Poulsen, C., Chambers, B., & D'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66*, 423-458.

Luellen, J. K., Shadish, W. R., & Clark, M. H. (2005). Propensity scores: An introduction and experimental test. *Evaluation Review, 29*, 530-558. doi:

Mallery, J. L. & Mallery, J. G. (1999). The American legacy of ability grouping: Tracking reconsidered. *Multicultural Education, 7*, 13-15.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*, 5-11.

Mickelson, R. (2001). Subverting swann: First and second-generation segregation in the Charlotte-Mecklenburg schools. *American Educational Research Journal, 38*, 215-252.

Millsap, R. E. & Maydeu-Olivares, A. (Eds.). (2009). *The Sage handbook of quantitative methods in psychology*. Thousand Oaks, CA: Sage.

Mosteller, F., Light, R. J., & Sachs, J. A. (1996). Sustained inquiry in education: Lessons from skill grouping and class size. *Harvard Educational Review, 66*, 797-842.

Oakes, J. (1987). Tracking in secondary schools: A contextual perspective. *Educational Psychologist, 22*, 129-153.

Oakes, J. (1992). Can tracking research inform practice? Technical, normative, and political considerations. *Educational Researcher, 21*, 12-21. Retrieved from http://www.jstor.org/stable/117720610.2307/1177206

Oakes, J. & Guiton, G. (1995). Matchmaking: The dynamics of high school tracking decisions. *American Educational Research Journal, 32*(1), 3-33.

Oakes, J. & Lipton, M. (1992). Detracking schools: Early lessons from the field. *Phi Delta Kappan, 73*, 448-454.

Oakes, J., Selvin, M., Karoly, L., & Guiton, G. (1992). *Educational matchmaking: Academic and vocational tracking in comprehensive high schools*. Macomb, IL: National Center for Research in Vocational Education.

Oakes, J., & Wells, A. S. (1998). Detracking for high student achievement. *Educational Leadership*.

Oakes, J., Wells, A. S., Jones, M., & Datnow, A. (1997). Detracking: The social construction of ability, cultural politics, and resistance to reform. *Teachers College Record, 98*, 482-508.

Office of the Superintendent of Public Instruction. (2012). *Washington State Assessment Coordinator's Manual*. Retrieved from http://www.k12.wa.us/TestAdministration/Instructions/

Onwuegbuzie, A. J. (2003). Expanding the framework of internal and external validity in quantitative research. *Research in School, 10*, 71-89.

Patton, C. (2010). *Conditions, processes, and consequences of tracking: A comparative case study of the structures, practices, and identities embedded in a lower track and an upper*

track English class (Doctoral dissertation). Available from ProQuest. (UMI Number: 3435894)

Paul, K. M. & Plucker, J. A. (2004). Two steps forward, one step back: Effect size reporting in gifted education research from 1995-2000. *Roeper Review, 26.*

Perrone, K. M., Wright, S. L., Ksiazak, T. M., Crane, A. L., & Vannatter, A. (2010). Looking back on lessons learned: Gifted adults reflect on their experiences in advanced classes. *Roeper Review, 32,* 127-139. http://dx.doi.org/10.1080/02783191003587918

Pierre-Louis, F. (2010). *A case study of Markdale high school's implementation of heterogeneously-grouped classes in English, mathematics, science and social studies* (Doctoral dissertation). Available from UMI Dissertation Publishing. (UMI: 3423413)

Plessy v. Ferguson, 163 U.S. 537 (1896).

Powell, A. G. (2001). Notes on the origin of meritocracy in American schooling. *History of Education Quarterly, 41,* 73-80.

Preckel, F., Gotz, T., & Frenzel, A. (2010). Ability grouping of gifted students: Effects on academic self-concept and boredom. *British Journal of Educational Psychology, 80,* 451-472.

Roberts v. City of Boston, 59 198 (Mass. 1850).

Robinson, J. P. (2008). Evidence of a differential effect of ability grouping on the reading achievement growth of language minority Hispanics. *Educational Evaluation and Policy Analysis, 30,* 141-180. doi: 10.3102/0162373708317742

Rogers, K. B. (1993). Grouping the gifted and talented: Questions and answers. *Roeper Review, 16.*

Rogers, K. B. (2007). Lessons learned about educating the gifted and talented: A synthesis of the research on educational practice. *Gifted Child Quarterly, 51*, 382-396. doi:10.1177/0016986207306324

Rothman, R. (2012). A common core of readiness. *Educational Leadership*, 11-15.

Rubin, B. C. (2003). Unpacking detracking: When progressive pedagogy meets students' social worlds. *American Educational Research Journal, 40*, 539-573. doi: 10.3102/00028312040002539

Rubin, B. C. (2006). Tracking and detracking: Debates, evidence, and best practices for a heterogeneous world. *Theory Into Practice, 45*, 4-14.

Rubin, B. C. & Noguera, P. A. (2004). Tracking detracting: Sorting through the dilemmas and possibilities of detracking in practice. *Equity & Excellence in Education, 37*, 92-101. doi: 10.1080/10665680490422142

Schofield, J. W. (2010). International evidence on ability grouping with curriculum differentiation and the achievement gap in secondary schools. *Teachers College Record, 112*, 1492-1528.

Shields, C. M. (2002). A comparison study of student attitudes and perceptions in homogeneous and heterogeneous classrooms. *Roeper Review, 24*. doi: 02783193

Slavin, R. E. (1991). Are cooperative learning and "untracking" harmful to the gifted? *Educational Leadership*, 68-71.

Slavin, R. E. (1995). Detracking and its detractors: Flawed evidence, flawed values. *Phi Delta Kappan, 77*, 220-222.

Spector, P. E. (1981). *Research Designs*. Newbury Park, CA: Sage Publications.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence* . New York: Cambridge University Press.

Sunderman, G. L. (2003). Federal-state relationships and the implementation of No Child Left Behind: First impressions. *Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.*

Sutton, A. & Soderstrom, I. (1999). Predicting elementary and secondary school achievement with school-related and demographic factors. *The Journal of Educational Research, 92*(6), 330-338.

Swiatek, M. A. & Benbow, C. P. (1991). Ten-year longitudinal follow-up of ability matched and unaccelerated gifted students. *Journal of Educational Psychology, 83*, 528-538.

Tanner, D. & Tanner, L. (2007). *Curriculum development: Theory into practice*. Upper Saddle River, NJ: Pearson.

Teachers College Record, (2010). 112(5), 1492-1528. Retrieved from http://www.tcrecord.org ID Number 15684.

Terman, L. M. (1916). *The measurement of intelligence*. Retrieved from GOOGLE ebookstore

Thurlow, M. L. (2012). *Common core state standards: The promise and peril for students with disabilities*. Retrieved from http://www.calstat.org/textAlt/SpEDge_eng/ 2012_Summer_Edge/

Tieso, C. L. (2003). Ability grouping is not just tracking anymore. *Roeper Review, 26.*

Tomlinson, C. A. (2006). An alternative to ability grouping. *Principal Leadership*, 6, 31-32.

Turney, A. H. (1931). Intelligence, motivation, and achievement. *Journal of Educational Psychology, 22*, 426-434. doi: 10.1037/h0072395

Turney, A. H. (1931). The status of ability grouping. *Educational Administration and Supervision, 17*, 21-42.

Tyack, D. B. (1974). *The one best system: A history of American urban education.* Cambridge, MA: Harvard University Press.

Van Tassel-Baska, J. (2005). Gifted programs and services: What are the nonnegotiables? *Theory into Practice, 44*, 90-97.

Van Tassel-Baska, J., Feng, A. X., Brown, E., Bracken, B., Stambaugh, T., French, H., ... Bai, W. (2008). A study of differentiated instructional change over 3 years. *Gifted Child quarterly, 52*, 297-312. Retrieved from http://dx.doi.org/10.1177/0016986208321809

Venkatakrishnan, H. & Wiliam, D. (2003). Tracking and mixed-ability grouping in secondary school mathematics classrooms: A case study. *British Educational Research Journal, 29*, 189-204.

Venzant, T. T. (2006). *Tracking segregation: Experiences of African American high school students in low and high track math and English classes* (Doctoral dissertation). Available from ProQuest. (UMI Number: 3223737)

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Webb, N. M., Shavelson, R. J., & Haertel, E. H. (2007). Reliability coefficents and generalizability theory. In *Handbook of statistics: Psychometrics* (pp. 81-124.). Amsterdam, The Netherlands: Elsevier.

Webber, S. (2010). *Understanding the delivery and impact of gifted education in America: A phenomenological study* (Doctoral dissertation). Available from ProQuest. (UMI Number: 3424333)

Wells, A. S. & Oakes, J. (1996). Potential pitfalls of systemic reform: Early lessons from research on detracking. *Sociology of Education, 69*, 135-143.

Wenglinski, H. (2004). Closing the racial achievement gap: The role of reforming instructional practices. *Education Policy Analysis Archives, 12*(64).

Wiliam, D. & Bartholomew, H. (2004). It's not which school but which set you're in that matters: The influence of ability grouping practices on student progress in mathematics. *British Educational Research Journal, 30*, 279-293.

Wing-yi Cheng, R., Lam, S.-F., & Chung-yan Chan, J. (2008). When high achievers and low achievers work in the same group: The roles of group heterogeneity and processes in project-based learning. *British Journal of Educational Psychology, 78*, 205.

Wiseman, A. W. (2010). The uses of evidence for educational policymaking: Global contexts and international trends. *Review of Research in Education, 34*, 1-24.

Wright, D. B. (2006). Comparing groups in a before-after design: When t test and ANCOVA produce different results. *British Journal of Educational Psychology, 76*, 663-675.

Xiang, Y., Dahlin, M., Cronin, J., Theaker, R., & Durant, S. (2011). *Do high flyers maintain their altitude?* [White paper]. Washington, D.C.: Thomas B. Fordham Institute.

Yonezawa, S. & Jones, M. (2006). Students' perspectives on tracking and detracking. *Theory Into Practice, 45*, 15-23.

Yonezawa, S., Wells, A. S., & Serna, I. (2002, Spring). Choosing tracks: "Freedom of choice" in Detracking Schools. *American Educational Research Journal, 39*, 37-67. Retrieved from http://links.jstor.org/sici?sici=0002-8312%28200221%2939%3A1%3C37%3ACT%22OCI%3E2.0.CO%3B2-Y

Appendix A

Chi square analysis

Cohort 0

Table A1

Case processing summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
ELL * School	388	100.0%	0	0.0%	388	100.0%
FRL * School	388	100.0%	0	0.0%	388	100.0%
Gender * School	388	100.0%	0	0.0%	388	100.0%
Ethnicity * School	388	100.0%	0	0.0%	388	100.0%
Sped * School	388	100.0%	0	0.0%	388	100.0%

Table A2

*ELL * School*

Crosstab

Count

		School		
		0	1	Total
ELL	0	243	145	388
Total		243	145	388

Chi-Square Tests

	Value
Pearson Chi-Square	.[a]
N of Valid Cases	388

a. No statistics are computed because ELL is a constant.

Table A3

*FRL * School*

Crosstab

Count

		School		Total
		0	1	
FRL	0	215	126	341
	1	28	19	47
Total		243	145	388

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.213[a]	1	.644		
Continuity Correction[b]	.091	1	.763		
Likelihood Ratio	.211	1	.646		
Fisher's Exact Test				.634	.378
Linear-by-Linear Association	.213	1	.645		
N of Valid Cases	388				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 17.56.

b. Computed only for a 2x2 table

Table A4

*Gender * School*

Crosstab

Count

		School		Total
		0	1	
Gender	0	103	63	166
	1	140	82	222
Total		243	145	388

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.042[a]	1	.838		
Continuity Correction[b]	.010	1	.922		
Likelihood Ratio	.042	1	.838		
Fisher's Exact Test				.916	.460
Linear-by-Linear Association	.042	1	.838		
N of Valid Cases	388				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 62.04.

b. Computed only for a 2x2 table

Table A5

*Ethnicity * School*

Crosstab

Count

		School		Total
		0	1	
Ethnicity	1	135	83	218
	2	77	40	117
	3	9	5	14
	4	8	6	14
	5	11	11	22
	6	3	0	3
Total		243	145	388

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4.038[a]	5	.544
Likelihood Ratio	5.010	5	.415
Linear-by-Linear Association	.110	1	.740
N of Valid Cases	388		

a. 2 cells (16.7%) have expected count less than 5. The minimum expected count is 1.12.

Table A6

*Special education * School*

Crosstab

Count

		School		Total
		0	1	
Sped	0	241	144	385
	1	2	1	3
Total		243	145	388

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.021[a]	1	.885		
Continuity Correction[b]	0.000	1	1.000		
Likelihood Ratio	.021	1	.884		
Fisher's Exact Test				1.000	.686
Linear-by-Linear Association	.021	1	.885		
N of Valid Cases	388				

a. 2 cells (50.0%) have expected count less than 5. The minimum expected count is 1.12.

b. Computed only for a 2x2 table

Cohort 1

Table A7

Case processing summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
ELL * School	475	100.0%	0	0.0%	475	100.0%
FRL * School	475	100.0%	0	0.0%	475	100.0%
Gender * School	475	100.0%	0	0.0%	475	100.0%
Ethnicity * School	475	100.0%	0	0.0%	475	100.0%
Sped * School	475	100.0%	0	0.0%	475	100.0%

Table A8

*ELL * School*

Crosstab

Count

		School		Total
		0	1	
ELL	0	295	180	475
Total		295	180	475

Chi-Square Tests

	Value
Pearson Chi-Square	.[a]
N of Valid Cases	475

a. No statistics are computed because ELL is a constant.

Table A9

*FRL * School*

Crosstab

Count

		School		Total
		0	1	
FRL	0	223	142	365
	1	72	38	110
Total		295	180	475

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.682[a]	1	.409		
Continuity Correction[b]	.510	1	.475		
Likelihood Ratio	.688	1	.407		
Fisher's Exact Test				.434	.238
Linear-by-Linear Association	.681	1	.409		
N of Valid Cases	475				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 41.68.

b. Computed only for a 2x2 table

Table A10

*Gender * School*

Crosstab

Count

		School		Total
		0	1	
Gender	0	142	72	214
	1	153	108	261
Total		295	180	475

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.989[a]	1	.084		
Continuity Correction[b]	2.669	1	.102		
Likelihood Ratio	3.001	1	.083		
Fisher's Exact Test				.088	.051
Linear-by-Linear Association	2.983	1	.084		
N of Valid Cases	475				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 81.09.

b. Computed only for a 2x2 table

Table A11

*Ethnicity * School*

Crosstab

Count

		School		Total
		0	1	
Ethnicity	1	152	108	260
	2	80	43	123
	3	20	12	32
	4	20	5	25
	5	21	11	32
	6	2	1	3
Total		295	180	475

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	5.516[a]	5	.356
Likelihood Ratio	5.842	5	.322
Linear-by-Linear Association	2.984	1	.084
N of Valid Cases	475		

a. 2 cells (16.7%) have expected count less than 5. The minimum expected count is 1.14.

Table A12

*Special education * School*

Crosstab

Count

		School		
		0	1	Total
Sped	0	294	179	473
	1	1	1	2
Total		295	180	475

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.125[a]	1	.724		
Continuity Correction[b]	0.000	1	1.000		
Likelihood Ratio	.121	1	.728		
Fisher's Exact Test				1.000	.615
Linear-by-Linear Association	.125	1	.724		
N of Valid Cases	475				

a. 2 cells (50.0%) have expected count less than 5. The minimum expected count is .76.

b. Computed only for a 2x2 table

Appendix B

Independent Sample t-tests

Cohort 0

Table B1

*Full-day absences * School*

Group Statistics

School		N	Mean	Std. Deviation	Std. Error Mean
Num_of_full_day_absences	0	240	12.29	12.081	.780
	1	143	13.27	12.265	1.026

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Num_of_full_day_absences	Equal variances assumed	.396	.530	-.764	381	.445	-.981	1.284	-3.505	1.543
	Equal variances not assumed			-.761	295.040	.447	-.981	1.288	-3.517	1.555

Cohort 1

Table B2

*Full-day absences * School*

Group Statistics

School		N	Mean	Std. Deviation	Std. Error Mean
Num_of_full_day_absences	0	295	13.29	12.393	.722
	1	180	14.25	11.493	.857

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Num_of_full_day_absences	Equal variances assumed	.470	.493	-.840	473	.401	-.958	1.141	-3.200	1.283
	Equal variances not assumed			-.856	400.378	.393	-.958	1.120	-3.160	1.243

Cohort 0

Table B3

*Grade 8 WASL * School*

Group Statistics

School		N	Mean	Std. Deviation	Std. Error Mean
Grade_8_WASL_Reading_2009	0	235	428.26	21.456	1.400
	1	140	424.37	17.528	1.481

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Grade_8_WASL_Reading_2009	Equal variances assumed	1.988	.159	1.811	373	.071	3.884	2.144	-.332	8.100
	Equal variances not assumed			1.906	337.959	.058	3.884	2.038	-.125	7.893

Cohort 1

Table B4

*Grade 8 MSP * School*

Group Statistics

School		N	Mean	Std. Deviation	Std. Error Mean
Grade_8_MSP_Reading_2010	0	276	419.99	20.731	1.248
	1	165	423.56	18.749	1.460

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Grade_8_MSP_Reading_2010	Equal variances assumed	4.180	.041	-1.812	439	.071	-3.568	1.969	-7.439	.302
	Equal variances not assumed			-1.858	372.641	.064	-3.568	1.920	-7.344	.208

Appendix C

Correlation Tables

Cohort 0

Table C1

Correlation table

Correlations

		School	ELL	FRL	Gender	Sped	Num_of_full_day absences	Ethnicity	Grade_8_WASL Reading_2009	Grade_10_HSPE Reading_2011	PSAT_Reading 2010	PSAT_Writing 2010	PSAT_Reading 2011	PSAT_Writing 2011
School	Pearson Correlation	1	a	.023	-.010	-.007	.039	.017	-.153**	.036	.034	.124	.089	.064
	Sig. (2-tailed)			.645	.839	.885	.445	.741	.003	.479	.841	.460	.111	.254
	N	388	388	388	388	388	383	388	375	384	38	38	321	321
ELL	Pearson Correlation	a	a	a	a	a	a	a	a	a	a	a	a	a
	Sig. (2-tailed)													
	N	388	388	388	388	388	383	388	375	384	38	38	321	321
FRL	Pearson Correlation	.023	a	1	-.014	-.033	.034	.130*	.009	-.133**	-.223	-.148	-.097	-.093
	Sig. (2-tailed)	.645			.780	.520	.512	.010	.862	.009	.178	.376	.081	.097
	N	388	388	388	388	388	383	388	375	384	38	38	321	321
Gender	Pearson Correlation	-.010	a	-.014	1	-.043	.164**	.009	.037	-.097	-.288	-.226	-.105	-.082
	Sig. (2-tailed)	.839		.780		.403	.001	.853	.477	.058	.079	.172	.061	.145
	N	388	388	388	388	388	383	388	375	384	38	38	321	321
Sped	Pearson Correlation	-.007	a	-.033	-.043	1	.222**	-.058	-.068	.006	a	a	-.037	-.059
	Sig. (2-tailed)	.885		.520	.403		.000	.258	.190	.911	0.000	0.000	.513	.295
	N	388	388	388	388	388	383	388	375	384	38	38	321	321
Num_of_full_day _absences	Pearson Correlation	.039	a	.034	.164**	.222**	1	.009	-.038	-.020	-.306	-.265	-.111*	-.143*
	Sig. (2-tailed)	.445		.512	.001	.000		.855	.463	.703	.065	.113	.049	.011
	N	383	383	383	383	383	383	383	370	379	37	37	316	316
Ethnicity	Pearson Correlation	.017	a	.130*	.009	-.058	.009	1	.006	-.067	-.132	-.200	-.014	-.103
	Sig. (2-tailed)	.741		.010	.853	.258	.855		.913	.192	.429	.228	.800	.065
	N	388	388	388	388	388	383	388	375	384	38	38	321	321
Grade_8_WASL_ Reading_2009	Pearson Correlation	-.153**	a	.009	.037	-.068	-.038	.006	1	.341**	.282	.303	.347**	.384**
	Sig. (2-tailed)	.003		.862	.477	.190	.463	.913		.000	.086	.064	.000	.000
	N	375	375	375	375	375	370	375	375	373	38	38	313	313
Grade_10_HSPE _Reading_2011_	Pearson Correlation	.036	a	-.133**	-.097	.006	-.020	-.067	.341**	1	.437**	.312	.528**	.436**
	Sig. (2-tailed)	.479		.009	.058	.911	.703	.192	.000		.006	.057	.000	.000
	N	384	384	384	384	384	379	384	373	384	38	38	320	320
PSAT_Reading_ 2010	Pearson Correlation	.034	a	-.223	-.288	a	-.306	-.132	.282	.437**	1	.697**	.700**	.533**
	Sig. (2-tailed)	.841		.178	.079	0.000	.065	.429	.086	.006		.000	.000	.001
	N	38	38	38	38	38	37	38	38	38	38	38	33	33
PSAT_Writing_20 10	Pearson Correlation	.124	a	-.148	-.226	a	-.265	-.200	.303	.312	.697**	1	.587**	.637**
	Sig. (2-tailed)	.460		.376	.172	0.000	.113	.228	.064	.057	.000		.000	.000
	N	38	38	38	38	38	37	38	38	38	38	38	33	33
PSAT_Reading_ 2011	Pearson Correlation	.089	a	-.097	-.105	-.037	-.111*	-.014	.347**	.528**	.700**	.587**	1	.735**
	Sig. (2-tailed)	.111		.081	.061	.513	.049	.800	.000	.000	.000	.000		.000
	N	321	321	321	321	321	316	321	313	320	33	33	321	321
PSAT_Writing_20 11	Pearson Correlation	.064	a	-.093	-.082	-.059	-.143*	-.103	.384**	.436**	.533**	.637**	.735**	1
	Sig. (2-tailed)	.254		.097	.145	.295	.011	.065	.000	.000	.001	.000	.000	
	N	321	321	321	321	321	316	321	313	320	33	33	321	321

** Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

a. Cannot be computed because at least one of the variables is constant.

Cohort 1

Table C2

Correlation table

Correlations

		School	ELL	FRL	Gender	Sped	Num_of_full_day absences	Ethnicity	Grade_8_MSP_Reading_2010	Grade_10_HSPE_Reading_2012	PSAT_Reading_2011	PSAT_Writing_2011	PSAT_Reading_2012	PSAT_Writing_2012
School	Pearson Correlation	1	.	-.038	.101	.016	.039	-.079	.198**	.246**	.178**	.176**	.265**	.236**
	Sig. (2-tailed)			.410	.027	.724	.401	.084	.000	.000	.000	.000	.000	.000
	N	475	475	475	475	475	475	475	441	469	453	453	235	235
ELL	Pearson Correlation	.	.	.	.	.	.	.	.	.	.	.	.	.
	Sig. (2-tailed)													
	N	475	475	475	475	475	475	475	441	469	453	453	235	235
FRL	Pearson Correlation	-.038	.	1	.057	.041	.084	.115*	-.086	-.169**	-.244**	-.251**	-.302**	-.265**
	Sig. (2-tailed)	.410			.218	.368	.068	.012	.071	.000	.000	.000	.000	.000
	N	475	475	475	475	475	475	475	441	469	453	453	235	235
Gender	Pearson Correlation	.101*	.	.057	1	-.008	.108*	.004	.198**	.061	.167**	.107*	.060	.116
	Sig. (2-tailed)	.027		.218		.869	.018	.937	.000	.185	.000	.022	.358	.077
	N	475	475	475	475	475	475	475	441	469	453	453	235	235
Sped	Pearson Correlation	.016	.	.041	-.008	1	-.028	-.045	-.025	.038	.019	-.004	.	.
	Sig. (2-tailed)	.724		.368	.869		.545	.327	.601	.418	.683	.938	0.000	0.000
	N	475	475	475	475	475	475	475	441	469	453	453	235	235
Num_of_full_day_absences	Pearson Correlation	.039	.	.084	.108*	-.028	1	-.034	-.028	-.023	-.047	-.089	-.039	-.065
	Sig. (2-tailed)	.401		.068	.018	.545		.465	.551	.619	.323	.057	.555	.319
	N	475	475	475	475	475	475	475	441	469	453	453	235	235
Ethnicity	Pearson Correlation	-.079	.	.115*	.004	-.045	-.034	1	.051	-.029	-.049	-.058	-.078	-.027
	Sig. (2-tailed)	.084		.012	.937	.327	.465		.281	.535	.297	.216	.234	.686
	N	475	475	475	475	475	475	475	441	469	453	453	235	235
Grade_8_MSP_Reading_2010	Pearson Correlation	.198**	.	-.086	.198**	-.025	-.028	.051	1	.345**	.432**	.431**	.368**	.370**
	Sig. (2-tailed)	.000		.071	.000	.601	.551	.281		.000	.000	.000	.000	.000
	N	441	441	441	441	441	441	441	441	435	420	420	215	215
Grade_10_HSPE_Reading_2012	Pearson Correlation	.246**	.	-.169**	.061	.038	-.023	-.029	.345**	1	.515**	.508**	.587**	.509**
	Sig. (2-tailed)	.000		.000	.185	.418	.619	.535	.000		.000	.000	.000	.000
	N	469	469	469	469	469	469	469	435	469	448	448	235	235
PSAT_Reading_2011	Pearson Correlation	.178**	.	-.244**	.167**	.019	-.047	-.049	.432**	.515**	1	.720**	.747**	.674**
	Sig. (2-tailed)	.000		.000	.000	.683	.323	.297	.000	.000		.000	.000	.000
	N	453	453	453	453	453	453	453	420	448	453	453	225	225
PSAT_Writing_2011	Pearson Correlation	.176**	.	-.251**	.107*	-.004	-.089	-.058	.431**	.508**	.720**	1	.712**	.730**
	Sig. (2-tailed)	.000		.000	.022	.938	.057	.216	.000	.000	.000		.000	.000
	N	453	453	453	453	453	453	453	420	448	453	453	225	225
PSAT_Reading_2012	Pearson Correlation	.265**	.	-.302**	.060	.	-.039	-.078	.368**	.587**	.747**	.712**	1	.787**
	Sig. (2-tailed)	.000		.000	.358	0.000	.555	.234	.000	.000	.000	.000		.000
	N	235	235	235	235	235	235	235	215	235	225	225	235	235
PSAT_Writing_2012	Pearson Correlation	.236**	.	-.265**	.116*	.	-.065	-.027	.370**	.509**	.674**	.730**	.787**	1
	Sig. (2-tailed)	.000		.000	.077	0.000	.319	.686	.000	.000	.000	.000	.000	
	N	235	235	235	235	235	235	235	215	235	225	225	235	235

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

a. Cannot be computed because at least one of the variables is constant.