5-7-2011

The Effects of Gifted Programming on Student Achievement: Differential Results by Race/Ethnicity and Income

Kelley M. Dean
Georgia State University

THE EFFECTS OF GIFTED PROGRAMMING ON STUDENT ACHIEVEMENT: DIFFERENTIAL RESULTS BY RACE/ETHNICITY AND INCOME

A Dissertation
Presented to
The Academic Faculty

By

Kelley M. Dean

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in
School of Public Policy

Georgia State University & Georgia Institute of Technology

May 2011

"The Effects of Gifted Programming on Student Achievement: Differential Results by Race/Ethnicity and Income"

Approved by:

Dr. David Sjoquist, Advisor
Andrew Young School of Policy Studies
Georgia State University

Dr. James Ainsworth
Sociology
School of Sociology
Georgia State University

Dr. Gary Henry
Carolina Institute for Public Policy
University of North Carolina-Chapel Hill

Dr. Gordon Kingsley
School of Public Policy
Georgia Institute of Technology

Dr. Mary Beth Walker
Andrew Young School of Policy Studies
Georgia State University

Date Approved: December 10, 2010

ACKNOWLEDGEMENTS

It is a pleasure to thank those who made this dissertation possible. My chair, David Sjoquist, I owe my deepest gratitude for his willingness to take on this thesis and remaining extremely responsive and supportive throughout the entire process. I owe many thanks to Gary Henry for inviting me to work with him when I first entered this graduate program. Through those research projects I learned a tremendous amount and developed this dissertation research. I am grateful for each member of my committee - Mary Beth Walker, Jim Ainsworth, and Gordon Kingsley - for offering their insight and knowledge to develop this work. I also greatly appreciate the continued help and support of Kevin Fortner, Dana Rickman, Grace Chikoto, and Nicholas Harvey. Finally, endless thank you's to my family and best friend for their love, prayers, and faith.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The central research question is the extent to which gifted programming effects student academic outcomes of gifted as compared to not-gifted students and how this differs by race/ethnicity and/or poverty status. Since the identification of elementary school students as gifted is not random, propensity score matching is used to remove this bias in the estimates of the effects. A matched sample of North Carolina middle school students based on individual level data of both gifted and not-gifted students of varied racial/ethnic groups and income levels is used for this analysis. This enables a comparison of sixth, seventh, and eighth grade student achievement to determine the extent to which participating in gifted programming differentiates effects by race/ethnicity and poverty status. I show the additional test score gain, if any, from being in gifted programming compared to students not participating in gifted programs. Variations in gifted program effects across race/ethnicity and income are assessed.

This research adds empirical evidence to the more qualitatively focused gifted debate by analyzing differences in student outcomes between gifted and not-gifted students in North Carolina. Since black and lower income students are less likely to participate in gifted programs, they disproportionately encounter less experienced teachers, lower expectations, and fewer resources. The extent to which these additional learning supports translate to differences in student outcomes are analyzed.

CHAPTER ONE: INTRODUCTION AND OVERVIEW

As a result of the No Child Left Behind Act passed in 2001, significant attention has been paid to underachieving students, often to the detriment of "gifted" learners. The requirement to disaggregate student data highlights achievement disparities among racial/ethnic groups, classes, and other subgroups (Barton, 2003; Ferguson, 2002; Peske & Haycock, 2006). However, there is much less intentional effort regarding the education of "gifted" students as teachers and principals often focus on low-performing students in order to meet requirements and secure funding.

Students in gifted programs are most often identified in elementary school and receive extra services including more advanced curriculum, additional resources, better teachers, and more challenging learning environments than their not-gifted peers. Critics of gifted programs purport that students who do not receive this label - and are therefore not provided the additional learning supports - are now at a disadvantage (Oakes, 1985; Smith-Maddox & Wheelock, 1995). Yet, proponents of gifted programs argue that additional services are needed for these "higher ability" students in order for them to reach their academic potential (Fiedler & Lange, 1993; Johnsen & VanTassel-Baska, 2006).

For my purposes, a student is characterized or labeled as "gifted" if he or she is formally identified as such through his/her school; this labeling is based on student assessments, most often test scores, and also relies on teacher perception. The treatment examined in this study is being categorized as gifted *and* being enrolled in the gifted program. The state of North Carolina mandates a three-step process of screening, identification, and placement into gifted programming (North Carolina Department of

Public Instruction, 2010). There is not however, a state-mandated cut-off for identification resulting in varying parameters within each local school district. Students in gifted programming receive additional resources including enhanced services and educational opportunities; therefore, middle school gifted programs are analyzed in order to describe and assess their impact on student outcomes.

It may be particularly useful to examine the potential differential effects of gifted programming across race/ethnicity and income. For instance, while examining social background and giftedness, Stamm (2007) finds that this label reproduces social hierarchies resulting in more negative effects for minority students. It is apparent from a significant volume of research that black and lower income students are much less likely to be labeled gifted (Delcourt, Cornell, & Goldberg, 2007; VanTassel-Baska, Feng, & Evans, 2007), and that the low levels of identification and lack of minority gifted students are clear, salient problems in research (Ford, Harris, Tyson, & Trotman, 2002; Morris, 2002; Ford & Grantham, 2003; Baldwin, 2005; Gallagher, 2005). There is much less known about the extent to which gifted programming impacts educational outcomes. The fact that minority and lower income students experience worse teachers than their more affluent peers (Clotfelter, Ladd, & Vigdor, 2006) further complicates the issue. As noted by Murnane and Steele (2007:15):

> Perhaps the most urgent problem facing American education…is the unequal distribution of high-quality teachers. Poor children and children of color are disproportionately assigned to teachers with the least preparation and the weakest academic backgrounds. Teacher turnover is high in schools that serve large shares of poor or nonwhite students because the work is difficult, and the teachers who undertake it are often the least equipped to succeed.

Two studies by Clotfelter, Ladd, and Vigor (2005, 2006), specifically using North Carolina data, find that the distribution of novice teachers disadvantages black students and that white and higher income students are much more likely to encounter higher quality teachers. These differences suggest that our nation's schools could be perpetuating existing disparities if in fact black and lower income students are disproportionately less likely to be labeled gifted and are systematically more likely to be exposed to weaker teachers. The potentially distinct outcomes in gifted and not-gifted students could be a result of differences in teacher quality, peers, and other additional learning resources afforded gifted students.

The central research question is the extent to which gifted *programming* effects student academic outcomes of gifted as compared to not-gifted students and how this differs by race/ethnicity and/or poverty status. Since the identification of elementary school students as gifted is not random, propensity score matching is used to remove this bias in the estimates of the effects. A matched sample of North Carolina middle school students based on individual level data of both gifted and not-gifted students of varied racial/ethnic groups and income levels is used for this analysis. This enables a comparison of sixth, seventh, and eighth grade student achievement to determine the extent to which participating in gifted programming differentiates effects by race/ethnicity and/or poverty status. I show the additional test score gain, if any, from being in gifted programming compared to students not participating in gifted programs. Variations in gifted program effects across race/ethnicity and income are assessed. I also include a measure in the empirical test that distinguishes the effects of participating in the gifted program from the effects of just being labeled as gifted in elementary school.

Note, however, that gifted identification and programming differ at the local level. One study using a national sample of students concludes that there is differential access to gifted programs and resources (Baker, 2001). Access to gifted programming therefore varies by race/ethnicity, socioeconomic level, locale, and region. Such differences by school also exist within states. I address this through the use of school characteristics within hierarchical linear models that control for such variations with schools at the third level.

I hypothesize that gifted students will have greater academic gains than not-gifted students; black and lower income gifted students will increase test scores at a higher rate than black and lower income not-gifted students. This is because gifted students are exposed to higher quality teachers, curriculum, and peers. This is especially important for black and lower income students who often experience the worst learning environments.

This work has important implications for understanding the extent to which our current system provides educational opportunities for traditionally underserved students. One must consider the implications of black and lower income students being less likely to receive the label and thus no gifted programming (Baldwin, 2005; Gallagher, 2005; Brown, Avery, VanTassel-Baska, Worley, & Stambaugh, 2006; Welner & Burris, 2006) as this often results in an even greater disadvantage in the educational system. Therefore, gifted programming is an issue of equity in which some students may receive less rigorous learning experiences.

This research quantifies the academic performance of a cohort of North Carolina middle school students and compares the growth of students who receive gifted programming to similar not-gifted students who had the propensity to be gifted. North Carolina is a state with racial/ethnic and economic diversity enabling an analysis of potential effects on a wide population of students. Several authors (Clotfelter, Ladd, & Vigdor, 2005, 2006, 2007; Henry, Fortner, & Thompson, 2010; Henry, Thompson, Fortner, Rickman, & Zulli-Lowe, 2008) have also conducted multiple education outcomes studies on the state offering a basis upon which I can build. This research therefore, adds empirical evidence to the more qualitatively focused gifted debate by analyzing differences in student outcomes between gifted and not-gifted students in North Carolina. Since black and lower income students are less likely to participate in gifted programs, they disproportionately encounter less experienced teachers, lower expectations, and fewer resources. The extent to which these additional learning supports translate to differences in student outcomes are analyzed.

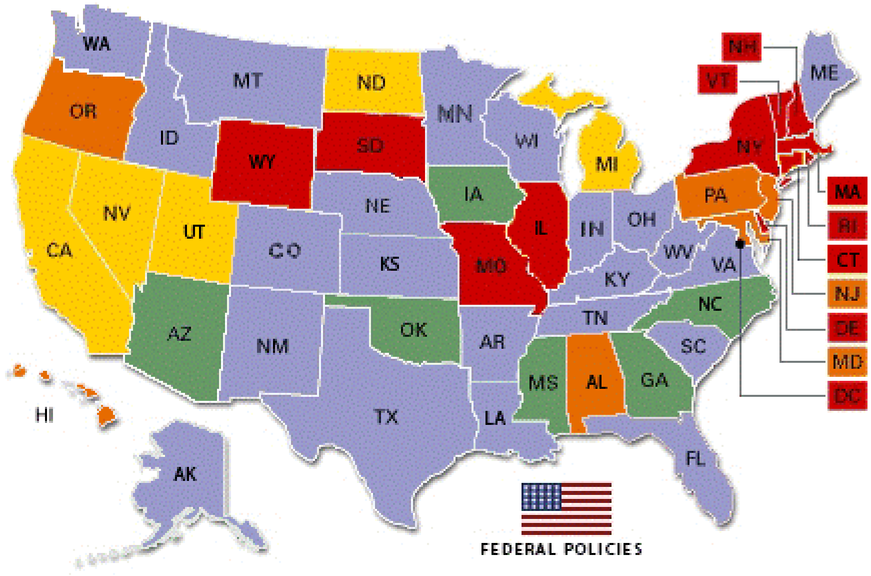
This dissertation is organized as follows. In Chapter 2, the gifted literature and relevant theories are discussed. Contributions of this research to the literature are also included. The third chapter details the research questions, hypotheses, data, and measures. The analytic approach is also explained including the basis for the final hierarchical linear math and reading models. The matched samples, descriptive results, and findings of the hierarchical linear models detailing effects of gifted programming are presented in chapter four. A key component of this research is an examination of differences in student outcomes for gifted compared to not-gifted middle school students for the state of North Carolina. The final chapter provides conclusions and policy implications.

CHAPTER TWO: LITERATURE REVIEW/THEORY

The gifted student literature has largely focused on the process of identifying gifted students and heavily utilizes qualitative studies. This analysis advances research in the field through a quantitative examination of the outcomes of students in gifted programming. These findings provide evidence to substantiate the theoretical arguments detailed below.

2.1 Gifted Education Overview – Ability Groups and Tracking

Gifted student programming remains a controversial issue in education today; the idea of labeling students and providing additional services to a select group of students is an area of much debate. Only twelve states have neither gifted funding nor a mandate for programming (Davidson Institute for Talent Development, 2010); although gifted education is extensive, there are widespread criticisms and biases. Gifted education programming and funding mandates are displayed graphically in Figure 1 below by state.



LEGEND
Gifted Programming is mandated: fully funded by state
Gifted programming is mandated: partially funded by state
Gifted programming is mandated: no gifted funding is available
Gifted Programming is not mandated; gifted funding is available
Gifted programming is not mandated; no gifted funding is available

Source: http://www.davidsongifted.org/db/StatePolicy.aspx

Figure 1 Gifted Education Policies

Opponents of gifted programs argue that students who are not identified as gifted and placed in the higher track, consisting of more advanced courses, are now disadvantaged; they are taught by less qualified teachers and develop lower educational aspirations (LeTendre, Hofer, & Shimizu, 2003; Lucas, 1999). LeTendre et al. (2003) examine tracking across three different places, Germany, Japan, and the United States to determine the extent to which the definition varies across cultures. Although tracking exists in each place, the authors conclude that the tracking process in the United States is much less based on accurate assessments and established systems, thus threatening equity and re-creating segregation. Based on Third International Math and Science Study (TIMSS) respondents, lower socioeconomic status (SES) schools are offered fewer courses and opportunities to learn than higher SES schools.

Other criticisms of the formal separation of students include the negative effects of lowered motivation and self-esteem on students who are not tracked (Oakes, 1985; Smith-Maddox & Wheelock, 1995). Oakes (1995) details the sorting process of students in 25 junior high and high schools across the country. She portrays some of the very negative experiences resulting from the tracking of students in lower ability tracks. Traditionally underserved students are often given the least in their educational experience. Smith-Maddox and Wheelock (1995) discuss the abolishment of tracking for students completely and the positive effect this may have on the classroom experiences of the most marginalized students. The role of counselors and educating both parents and students on the significance of high school course selection are highlighted. Therefore instead of separating students according to perceived ability, mixed grouping could result in leadership development of “high ability” students and enrichment to “low ability” students. It is possible that gifted and not-gifted students learning together in heterogeneous groupings might negate the potentially negative effects of not participating in gifted programming.

There is also evidence of benefits to additional programming for students identified as gifted. Proponents of gifted education contend that some children need additional enrichment in order to remain engaged in school and be challenged to their full potential (Fiedler & Lange, 1993; Johnsen & VanTassel-Baska, 2006). Several studies examine potential differences in learning based on the grouping of students. Findings include that high ability students experience greater gains when grouped with other high ability students (Kulik & Kulik, 1992) and need to be together in order to "thrive" academically (Gessner, 2008).

Kulik & Kulik (1992) conduct a meta-analysis of existing grouping programs for students. These studies focus on five distinct instructional programs that group students by ability: multilevel classes, cross-grade programs, within-class grouping, enriched classes for gifted students, and accelerated classes. Student gains increase in response to program type; the more curriculum changes in place, the more students gain. Multilevel classes, where students are grouped into separate classes based on ability but all cover the same curriculum, have no little or no effect on student outcomes as there are few curriculum adjustments made. Cross-grade and within-class programs result in small, positive effects (0.30 and 0.25 standard deviations, respectively) since teachers divide students into ability groups and then guide curriculum accordingly. Finally, enrichment and acceleration programs that entail the most curriculum changes have the largest effects on student achievement (0.41 and 0.87 when compared to students of the same age and -0.02 when compared to older students). Therefore, Kulik and Kulik (1992) conclude that there is benefit to ability grouping and find the most benefit is experienced by those in high ability groups. Yet, they do not find that lower ability groups are harmed in the process. The improved performance is attributed to the curriculum differentiation that occurs in more advanced classes as opposed to grade-level focus of regular courses.

Similarly, Gessner (2008) finds that students learn at different rates and argues those who are high ability need more educational opportunities to meet their needs. He purports that tracking systems create the higher ability groups that gifted students need in order to “thrive.” He finds that ability grouping is being replaced by cooperative learning and differentiation which are often challenging for teachers to implement and proven largely unsuccessful. This supports the findings of Kulik and Kulik (1992) that learning environments need to be enriched and more challenging for high ability students. Gessner (2008) also criticizes No Child Left Behind legislation as limiting the more advanced learning experiences of gifted students.

Note that there are mixed findings of the effect of heterogeneous classrooms on both “low” and “high ability” students. While gifted students may benefit from learning together, there is evidence that ability grouping and/or tracking; increases learning for low ability students while decreasing learning for high ability students (Argys, Rees, & Brewer, 1996); results in no difference between groups (Betts & Shkolnik, 2000); and negatively affects higher ability students (Loveless, 1999; Loveless & Thomas, 2009).

Arygs et al. (1996) examine a nationally representative sample of students through a National Center for Education Statistics (NCES) survey. An education production function is used to estimate the impact of tracking on academic outcomes while controlling for various factors including class size and teacher education. Four test score equations are used: high ability, average ability, low ability, and heterogeneous. They find that ending tracking would largely increase performance of lower track students; however, the loss of tracking would negatively impact higher track classes. Also using national survey data, Betts & Shkolnik (2000) argue that the impact of ability grouping on student performance is not as strong as previous studies conclude. In this study principals are asked about ability grouping or tracking in their schools. The authors contend there may be a difference between formal and more informal ability grouping and that the latter would not be captured in survey data. They argue strong conclusions of differential effects of ability grouping are not yet not warranted based on existing research. Finally a more recent study examines tracking and middle school student achievement in Massachusetts (Loveless & Thomas, 2009). While there is no significant difference in student performance between schools with tracked and detracked English Language arts programs, there are highly significant differences in performance across math tracks. Tracked schools have significantly more students scoring at the advanced and proficient level than untracked schools in math; performance increases as the number of tracks increase. The opposite is true of detracked schools; there are more failing students as the number of tracks decreases.

While there are both positive and negative effects of ability grouping and tracking students found in the literature, the benefit or detriment to all students resulting from these techniques are inconclusive.

2.2 Long-Term Effects of Gifted Programming

Gifted programming also has potential long-term educational effects, since it is generally expected to lead to these students' placement on high tracks and obtaining greater student outcomes. Gifted students are more likely to take advanced courses, have higher educational attainment, and experience social benefits. Brody and Mills (2005) synthesize findings of numerous other reports including the work of the Center for Talented Youth at John Hopkins University and several talent search programs. They detail the numerous benefits found in other research including academic, social, and familial. Benefits are also found for under-represented students who participate in the talent search program; these students most often experience growth in maturity and independence in addition to thinking skills and create more possibilities for their future.

Since high ability students are more likely to achieve in school, the disproportionate number of white students in these programs by race is problematic (Baldwin, 2005; Brown et al., 2006). The tracking created often results in both racially and socially stratified learning environments in which not-gifted students suffer; some contend there is a clear lack of learning in the low track. Welner & Burris (2006) use data from two case studies, one in Long Island focused on “winning them over” and the other near Pittsburgh, Pennsylvania attempting to “take them on” to explore detracking attempts. The approach to detracking depends on the school environment. Students in the first high school are an example of “winning them over” as classes are gradually detracked, eventually resulting in positive improvement for all students including those formerly resigned to the lowest track. Parents of students in lower tracks of the second high school “take them on” by organizing and fighting the tracking process that is clearly stratified by race/ethnicity and income. Eventually a court decision orders the school to detrack, however this process is much more political and resistance is apparent. The authors further discuss the potential risk of what occurs after the court order ends and decisions are returned to school administrators. They note how easy it is to ignore parents of lower track students when they are politically invisible. This supports the argument that this politically charged issue of tracking results in class reproduction of students essentially relegating black and lower income students to failure. Bowles & Gintis (2002) contend that capitalism creates significant inequality that is transmitted across generations and results in sharp disparities according to income levels; schools actually reinforce rather than correct this inequality. Therefore, gifted labeling is claimed to both re-segregate schools and increase the test score gap between white and black students.

The distinct test score gap in this country as measured by both race/ethnicity and poverty status is a critical concern in education today. White and high SES students score significantly higher than black and low SES students on standardized tests (Haycock, 2001). Thus poverty is another potential influence on disparities in gifted student outcomes. Stiefel, Schwartz, and Ellen (2007) find that differences in test scores remain after controlling for school and individual poverty. Contrary to the argument that poverty accounts for racial/ethnic differences, this study finds that race/ethnicity affects student performance above and beyond income level. OLS regressions and hierarchical linear models are used that include individual, program specific (disability/special education, limited English proficient, and gifted and talented), and school specific variables for one year of fifth grade and eighth grade students in New York City public schools. Racial test score gaps across and within schools are compared. Much of the racial gap is determined by previous student performance which is lower for black and Hispanic students. Gaps also remain when classroom characteristics are controlled.

If white students and higher income are indeed more likely to be in the gifted program with more challenging courses, higher expectations, and additional resources, this gap in student performance may be exacerbated.

> The vast majority of public school students in the United States are tested, ranked, and segregated into separate ability groups and classes based on standardized test performance. Although schools have been tracking students since the 1920s, the practice has always rested uneasily with the principal of economic opportunity. Critics, point out that low-income and minority students are placed disproportionately in the low-tracks, argue that tracking unnecessarily perpetuates inequality. Proponents argue that tracking is a necessary response to

> the widely divergent abilities and needs of students. ("Teaching Inequality," 1989)

Although written over twenty years ago, these findings still ring true today. It is currently argued that tracking is the newest form of segregation (Blackburn, 1999; Gallagher, 2005). As forced integration and affirmative action programs have all been struck down, grouping students by perceived ability is arguably the latest attempt to maintain the privilege of some students while continuing to deny others. Stiefel, Schwartz, and Chelmman (2007) attempt to study low achieving students in New York state elementary and middle schools. They find schools are so highly segregated that over half are too homogenous to compare academic performance by race/ethnicity. Achievement gaps are actually greater across schools than within schools; homogeneity leads to accountability as test results cannot be disaggregated. Similarly Orfield, Frankenburg, & Garces (2008) summarize the responses of 553 social scientists on voluntarily adopted racial integration plans by two school districts and conclude that race-conscious policies are still needed to maintain racial integration in schools.

2.4 Racial Matching of Teachers and Students

White and higher income students are usually perceived to have higher abilities than black and lower income students. This adds a racial component to gifted programming that lends further consideration. There are an increasing number of assessments used to determine whether or not a child is gifted, however these evaluations occur only *after* a student is recommended for consideration by their teachers. Thus teachers, and to a lesser extent parents, play a crucial role in this initial identification of gifted students. Individual perceptions and biases come into play and often result in the disproportionate identification of minority and low-income children.

Perceptions and racial influences are considered more generally in public administration literature. There is considerable work around the idea of racial matching and representative bureaucracies. Evidence exists that racially representative bureaucracies serve minority clientele more effectively than non-diverse organizations. Meier (1993) studies the role of Latino administrators as active representatives for Latino students. Using longitudinal data from twelve Florida school districts, Meier finds support for the hypothesis that street level bureaucrats are more likely than upper level bureaucrats to serves as active representatives. He also argues a critical mass of Latino administrators are needed in order to expect active representation of Latino students. As supported in other studies (Meier, 1975; Mosher, 1982; Selden, 1997), the central idea to representative bureaucracies is that people with similar characteristics, such as race/ethnicity, share values and beliefs that are better represented by people who share the similar characteristic or are reflective of the group.

Within education studies, the representation of persons within schools has been found to effect test scores (Meir, Wrinkle, & Polinard, 1999) and high school drop outs (Pitts, 2005). Meir et al. (1999) discuss the influence of representative bureaucracies on minorities compared to nonminorities in schools. The authors conclude there is not in fact a zero sum game for participants, but that both groups perform better with the presence of representative bureaucracies; their inclusion is more effective. Pitts (2005) addresses similar relationships (diversity, representation, and performance) using a racial mix of teachers, administrators, and students within Texas school districts. Significant relationships are found between teacher diversity and performance, but not manager diversity, again suggesting street level bureaucrats have a greater impact on student performance than management. The use of policy tools is influenced by the race/ethnicity of bureaucrats and the social construction they have of the students with whom they are interacting (Schneider & Ingram, 1993; Pitts, 2007). Schneider & Ingram (1993) purport that the social construction of target populations is a political component that needs to be addressed in research. They drive political scientists to consider this new dimension in their policy research. More recently, Pitts (2007) argues that ethnic representation is critical to considering the changing demographics across the country. He finds that ethnic representatives actually benefit the organization as a whole, not just target populations.

Economics literature also offers insight on the perceptions and influence of race in educational outcomes. One study (Bishop, Dudley, Mihaly, & Murphy, 2005) finds that black students with teachers of the same race score higher on standardized tests and are more likely to take college entrance exams. White males are also more likely to take college entrance exams when they have same race teachers. This study uses data from the Tennessee Project Student Teacher Achievement Ratios (STAR) that created a randomized experiment to examine the impact of class size on student achievement. In the original study, Dee (2004) finds a significant impact of a teacher of the same race on student performance in one year of analysis that is supported by Bishop et al. (2005) in their longitudinal study for the achievement of black students.

The idea of racial matching is also explored within gifted literature. Elhoweris, Mutua, Alsheikh, and Holloway (2005) conduct a study with elementary school teachers and find that when teachers are randomly assigned to groups, race/ethnicity is a factor in the referral decision to gifted programming. Other work also concludes that race/ethnicity matters in teacher's perceptions; in a national study, the classroom behavior of students was rated higher by black teachers than by white teachers, suggesting a potential bias or differential expectation of white teachers regarding black students (Downey & Pribesh, 2004). The authors also discuss the influence of cultural matching between teachers and students on student behavior. Such perceptions and biases of teachers are significant and directly contribute to learning experiences of students.

The lack of black and lower income children identified as gifted may be a result of differential teacher expectations. This is crucial because gifted programs often segregate schools by race/ethnicity and income, commonly resulting in very different experiences for more privileged children compared to traditionally underserved children. Students can be set on specific, almost pre-determined educational paths based on their perceived ability level.

2.5 Measures of Giftedness

In addition to the potentially negative impacts of disproportionate gifted programming and teacher perceptions, another prevalent issue is the evaluation of these students after they are recommended to the program or identified as potentially gifted. Many contend there is not an adequate measure of gifted ability. "The gifted field has too few technically sound screening instruments. The ubiquitous IQ test is almost always routinely used" (Pfeiffer & Jarosewich, 2007). Previously, gifted identification was solely based on standardized tests; however, lawsuits and changing views of education are beginning to lead to the use of multiple measures. Educators now acknowledge the idea of multiple intelligences (Gardner & Hatch, 1989) and attempt to assess these various ways of learning (Fiedler, Lange, & Winebrenner, 2002).

In one of the six myths these authors dispel, Fiedler et al. (2002) argue that ability grouping does not inherently discriminate against racial/ethnic minorities. Instead they argue that the use of inequitable assessment procedures have resulted in the under-identification of some students for gifted classes. "Wide-spread efforts are being made to overcome the inequities of overreliance on standardized test score data and assumptions that too often have been made about students who, although gifted, may not fit the stereotype of high achievers with positive attitudes toward school. The direction is away from sole reliance on standardized tests and toward improved approaches that include studying the behaviors of students for indicators that gifted potential exists" (Fiedler et al., 2002:110). The North Carolina Department of Public Instruction (DPI) is an example; DPI conveyed its desire "to minimize the role of psychometric information in the identification of students for placement" (as cited in George & Harrison, 2001: 10) with the implementation of revised policy recommendations.

Thus, gifted identification is increasingly determined by more inclusive measures including tests (standardized and other assessments), previous academic performance, teacher observations, and parent/student referrals due to numerous concerns with traditional identification processes. These new measures are a significant move away from the traditional use of criteria with proven test biases and inaccurate measures of students' potential (Roscigno & Ainsworth-Darnell, 1998; Boaler, 2003). "The IQ score has traditionally been the method of sorting out who is gifted and who might not be gifted. It is this tradition that has been called into question by educational researchers and scholars, due to the mounting evidence that IQ scores are not the only indicators of giftedness" (Baldwin, 2005: 105). Yet, there are still disproportionately few black and lower income students identified for these programs.

Although the types of assessments are increasing, students in most districts must be referred to the gifted program before any type of assessment takes place (McBee, 2006); the potential positive effects of multiple assessments leading to greater student identification may be not be realized if black and lower income students are not first recommended. In a study of all elementary schools across the state of Georgia, McBee (2006) uses gifted nomination status and gifted identification status to examine the gifted process across race/ethnicity and poverty status. He finds that Asian and White students are much more likely to be nominated than Black or Hispanic students. Higher income students, those not eligible for free or reduced lunch, are also more likely to be nominated. He concludes that inequalities in nomination, not assessment, may be the main reason for the underrepresentation of minority and lower income students in gifted programs.

Baldwin (2005) also attempts to better understand the lack of minority and lower income students in gifted programming. This issue is especially salient, he states, because of the increasing number of minority students in public schools. He warns against continuing to use current practices to identify students and presents more creative assessment techniques. The disproportionately high levels of labeling and placement of white and more affluent students into gifted programming may negatively influence the most traditionally underserved children.

2.6 Isolation Effect of Gifted Students

Tyson, Castellino, and Darity (2005) discuss the idea of isolation effects through a "disparity index" in a study on the "burden of acting white" in schools. The idea of this "burden" was first introduced by Ogbu and colleagues (Fordham & Ogbu, 1986; Gibson & Ogbu, 1991) and was largely accepted for at least a decade. The "burden of acting white" suggests there is an oppositional culture resulting in disparities in academic outcomes for historically oppressed groups (e.g. black students). These involuntary minorities resist the dominant group by opposing school goals.

More recently, both quantitative and qualitative studies are challenging aspects of this sociological theory (Ainsworth-Darnell & Downey, 1998; Carter, 2005; Downey & Ainsworth-Darnell, 2002; Lewis, 2003) through findings of different views across racial/ethnic groups and how they view occupational opportunity in addition to school goals. While examining gifted students, this "burden" is found to only exist in diverse schools where the proportion of black/lower income students in gifted/higher level classrooms do not match the percentage of black/lower income students in the overall school population (Tyson et al., 2005). There is an isolation effect or level of discomfort experienced by black students who are often one of few students identified as gifted in heterogeneous schools. These authors argue that the burden extends beyond race to income as well. Lower income students who manage to make it to the higher level courses are often shunned as well by their less affluent peers because they are deemed as privileged.

Findings from this study are also used to support the idea of an "anointment effect" stemming from the gifted label. "This suggests a sort of 'anointment effect,' whereby students are identified as high-achieving in the early grades, incorporate this 'anointment' into their self-perceptions, and act on these positive self-perceptions by continuing to achieve" (Sadowski, 2006 para. 20). Therefore, students identified in the early grades for gifted and talented programs incorporate the "anointment" in into how they view themselves potentially positively impacting their future achievement for years to come.

2.7 Gifted Effects Literature

Although limited, a small amount of research exists on the effects of gifted programming. Cornell, Delcourt, Goldberg, and Bland (1995) examine both achievement and self-concept of gifted and not-gifted students at the elementary school level. There are no teacher or classroom characteristics and limited student information (due to missing data) included in their analysis of variance. They find that minority gifted students score higher than minority students not in the gifted program; however, minority, gifted students score below their white gifted peers although still above grade level. The authors note the great extent of literature on the identification of gifted students, yet there is still a lack of studies on minority students that are placed in gifted programs. They recommend that future studies "investigate whether standardized test scores are equally predictive of academic success for both minority- and majority-group students" and also include analysis of socioeconomic variables (Cornell et al., 1995).

A 2007 study by Delcourt, Cornell, and Goldberg is the most related prior research examining gifted programming effects. Elementary school students in fourteen districts from ten different states are used in this analysis of 460 African-American and Caucasian/non-Hispanic second and third grade students. The authors classify gifted programs as special school, separate class, pull-out, and within class programs. High achieving students in districts without gifted programs and not-gifted students in "regular" classrooms were also included as comparison groups. Both cognitive and affective outcomes, including achievement tests and two student surveys (self-perception and motivation) from two points in time, are examined. Outcomes of students from traditionally underserved populations are a central aspect. Thus, characteristics of minority and lower income students are key subgroups of evaluation. Control variables include test scores upon entrance to one of the four program types and an index of economic disadvantage; the Hollingshead 4-factor index encompassing sex, and parental marital status, education, and profession.

Through an analysis of covariance, Delcourt et al. (2007) find that cognitive and affective student outcomes do vary across program type. Examining cognitive outcomes by race/ethnicity, white students both have higher achievement scores than black students on all five achievement subtests. The adjusted means for white students were larger than that of black students across math, reading, science, and social studies scales; differences in means ranged from four to ten points between white and black students. The test scores of black gifted students do remain above mean for their grade level and follow an upward trend from the fall of the first year to spring of the next, but black gifted students perform below their white gifted classmates.

However, student cognitive performance does not vary by race/ethnicity according to the type of gifted programming. Learning outcomes are significantly different across program type, but do not vary significantly across racial/ethnic groups, suggesting that ability grouping is an effective practice. Students in special school, separate class, and pullout programs perform higher than gifted students who are not in a gifted program and higher than most within-class program and not-gifted students. Math achievement is the exception with not-gifted students scoring significantly higher than gifted students not in a program and those in within-class programs.

The authors note that gifted students in within-class programs perform the lowest in all subjects when compared to their gifted peers in special schools, separate class programs, and pull out programs. As in other studies, an emphasis on differentiating curriculum for gifted learners is highlighted as greatly influencing the educational outcomes of these students. The authors contend gifted programming is valid and urge educators to improve evaluation practices to make sure all students' needs are continuously monitored and addressed.

Ford, Grantham, and Whiting (2008) offer a descriptive, exploratory study that examines peer pressure faced by black, gifted students. They discuss the limited research on the achievement gap related to high-ability students; "to our knowledge, no studies have examined the gap with Black students formally identified as gifted" (218). Their findings support existing research that black males are likely to underachieve due to negative views of intelligence, the "acting white" phenomenon, and stereotype threat.

2.8 Gaps in Existing Literature, sample limitations, methodological limitations

There are several gaps in existing gifted literature and weaknesses of existing studies. Perhaps the most apparent is a gap in the literature of the effects of gifted programming. It is clear that although much is known about the identification of gifted students, considerably less research exists on the actual impact of their participation in gifted programs (Robinson, 1990; Shore & Delcourt, 1996; Robinson & Clinkenbeard, 1998). For instance, a five state study on gifted educated programs concludes that "more individual state studies of policy impacts as well as program services need to be conducted" (Brown, Avery, VanTassel-Baska, Worley, & Stambaugh, 2006). National Association for Gifted Children program standards are examined to compare gifted education programs and policies. The authors find varying degrees of gifted education policies and note an emphasis on program identification over program development and teacher preparation. The authors recommend connecting education policies impacting gifted learners such as content standards, state assessments, and secondary courses including Advanced Placement, International Baccalaureate, and dual enrollment programs to measure student outcomes as technology improves accessibility to such data. After surveying three leading journals in gifted education, Delcourt et al. (2007:361) find that "surprisingly few studies have directly examined how students change over time after entering a gifted program." There is a clear gap in the literature regarding impacts of gifted student programming on academic outcomes.

While there are numerous studies examining differences in student performance of white students compared to black students, such studies are lacking in the area of gifted research. Ford et al. (2008) conclude that further information is needed on the academic performance of students in gifted programming to determine to what extent gaps in achievement occur and also to consider development of strategies to promote student achievement, particularly of underachievers.

There are also weaknesses of existing studies including the data and methods of analysis used. Several studies conduct meta-analyses of previous research often resulting in generalizations and outdated data. There are also few longitudinal students of students over time as most studies analyze two points in time. The Cornell et al. (1995) study uses an analysis of variance with a small sample of students. Delcourt et al. (2007) has similar weaknesses. In addition to a small sample of 460 students and simple method of analysis of covariance, fourteen different school districts in ten states are used. Students are also not necessarily comparable as various program types are defined that may or may not translate across the data from multiple states.

Finally, there are more conceptual concerns as well. There are potential problems of using survey data to identify the presence of ability groups and tracks as respondents may define these aspects in different ways. Several weakness are pointed out by Betts and Shkolnik (2000) including an unclear differentiation between ability grouping and curriculum tracking, lack of information regarding within classroom practices, distinction between degrees of ability grouping/tracking, and that group placement may in fact be an imperfect proxy of measured student ability. There is much room to improve the methods and quality of data that were used in existing studies.

2.9 Contributions to the Literature

This research addresses an understudied area by seeking to understand the effects of gifted programming on student achievement and how this varies by race/ethnicity and income. There is a clear gap in the literature on such quantitative effects with clearly defined gifted and not-gifted groups. The data used in this research enables a longitudinal examination of a cohort of similar students over their middle school years of sixth, seventh, and eighth grade. This research adds empirical evidence to the literature on gifted education as gifted programming effects are quantified through a matched sample providing less biased estimates of treatment effects than previous studies.

There is a large body of theoretical and qualitative work on identification of gifted students, yet quantitative work on the potential impact of gifted programming is lacking. This analysis takes the research a step further by focusing on academic outcomes after students have been placed in gifted programs.

The implications of the racial/ethnic and income disparities of gifted student programs are potentially very significant; life trajectories may be raised for some children while lowered for others. Analyzing the gifted programming some students receive is one way to develop an evidence based understanding of this issue and the many children it may affect. The extent to which gifted programming leads to the sorting of students and segregation of students in racially mixed schools is examined. If this perceived ability grouping results in the perpetuation of dividing students based on race/ethnicity and/or poverty status in schools, there are huge societal consequences for this structural mechanism. The damaging effect of being labeled not-gifted and therefore not receiving enhanced resources or programming may be especially problematic for black and lower income students who are disproportionately less likely to be identified as gifted (Baldwin, 2005; Gallagher, 2005; Brown, Avery, VanTassel-Baska, Worley, & Stambaugh, 2006; Welner & Burris, 2006).

The methods used also add to existing gifted effects literature. This work strengthens the previous Delcourt et al. (2007) research by using a much larger sample size and students across multiple years from a single state. Potential variation across programs is decreased since gifted programs within the same state are likely to be more comparable than gifted programs across numerous states. The analytical method is also improved from an analysis of covariance to hierarchical linear growth models; they more fully explain variation in the estimates and enable an analysis of change over time.

This study also uses a more highly specified, longitudinal model of a statewide cohort of middle school students. Four points in time are used to create matched samples and run growth models analyses examining academic outcomes of gifted and not-gifted students. This contributes to the literature by examining the relationship between race/ethnicity and income on student achievement, particularly gifted student achievement. Differences in academic performance of students identified as gifted are analyzed including the extent to which a gap exists by race/ethnicity or income among these "high ability" students.

Hierarchical linear models are used in gifted literature to examine test anxiety effects (Goetz, Preckel, Zeidner, & Schleyer, 2008; Ma & Wilkins, 2002; Preckel, Zeidner, Goetz, & Zeidner, 2008); however, HLM is not a method widely used in analysis of gifted academic outcomes. Therefore, the application of a matched sample of gifted and not-gifted students to a multi-level model is another contribution of this research. The use of varying slopes will provide an in-depth analysis of the potentially differential effects of race and income.

Understanding if gaps in achievement by race/ethnicity persist for gifted students offers a different lens through which to view education. The well researched differences in standardized test scores of the general population may also apply to gifted students. The extent to which students perceived to have higher ability vary in academic performance by race/ethnicity and income could greatly inform school practices and educational policies.

CHAPTER THREE: RESEARCH QUESTIONS AND METHODOLOGY

This study systematically analyzes differences in student performance by participation in gifted programs, race/ethnicity, and poverty status for middle school sixth, seventh, and eighth grade students. Research questions and hypotheses, sample and data, measures, and the analytic approach are detailed.

3.1 Research Questions and Hypotheses

This study seeks to quantify potential impact of gifted programming on middle school students, and how that differs by race/ethnicity and income. The central research question is the extent to which gifted programming effects student academic outcomes of gifted as compared to not-gifted students and how this differs by race/ethnicity and/or poverty status. Descriptive data is used to detail actual student outcomes and gains over time for a cohort of middle school students. Growth models are then used to predict the additional test score gain, if any, from being in gifted programming compared to similar students not participating in gifted programs. Variations in gifted program effects across race/ethnicity and income are assessed. I also include a measure in the empirical test that distinguishes the effects of participating in the gifted program from the effects of just being labeled as gifted in elementary school.

Research Question:

To what extent do student outcomes differ between gifted and not-gifted students, particularly by race/ethnicity and income:

- Who is enrolled in gifted programming in North Carolina middle schools and how does this vary by race/ethnicity or income?
- How do math and reading outcomes of gifted and not-gifted students compare throughout middle school? Does this vary by race/ethnicity and/or class?
- How do the predicted test scores and gains over time of gifted and not-gifted middle students compare in reading and math? Is this differentiated by race/ethnicity or income?

White gifted students and higher income gifted students are expected to outperform white not-gifted students and higher income not-gifted students. Black gifted students and lower income gifted students are also expected to perform higher than black not-gifted students and lower income not-gifted students. This is based on previous research that shows gifted students are more likely to take advance courses and have higher educational attainment than not-gifted students (Brody & Mills, 2005). Gifted students are more likely to experience better teachers and enriched educational programming; thus, it is expected that these students excel since they are challenged and developed in many more ways than not-gifted students (VanTassel-Baska, 2005). "To the extent that opportunities to learn are structured by schooling, then the path of coursework a student takes will sharply influence their human capital accumulation.

Knowledge acquired is cumulative" (Darity et al., 2001). This learning or lack of learning directly affects performance on standardized tests; students, no matter their capacity, are not going to do well on a test if they have not been exposed to the material (Hallinan and Sorenson, 1977). School environments and courses taken by students very much influence both learning experiences and educational outcomes (Bryk, Nagaoko, & Newmann, 2000).

The educational environment of gifted students is also potentially enriched due to peer learning effects. Students themselves help create the learning environment thus peer effects of gifted students grouped together is likely to result in an enhanced learning experience. Previous studies include comparisons of individual students in relation to their peers in education production functions (Hanushek, 1979; Hanushek, Kain, Markman, & Rivkin, 2003). Peer effects are commonly found to have a positive, significant effect on student outcomes; higher ability students positively impact the learning of their peers.

When comparing these groups across race/ethnicity and income, gifted black students and gifted lower income students are expected to gain more than both white and higher income students. Research shows that teacher education levels may matter for student outcomes (Darling-Hammond, 2007; Goldhaber & Brewer, 1997; Goldhaber & Brewer, 2000) and more specifically the combination of a teacher with an advanced degree in the subject taught contributes to student learning (Hawk, Coble, & Swanson, 1985). Since more experienced and in-field teachers tend to teach higher level students (Clotfelter et al., 2006; Murnane & Steele, 2007), it is expected that gifted students will benefit from this exposure. Additionally since lower income and minority students often encounter the least experienced teachers, when these students are in gifted programming the benefit to them is expected to be the greatest as they are experiencing a much more enriching learning environment than they would have otherwise. Therefore, the gains of gifted students are expected to be higher and also stratified by race/ethnicity and income.

There also may be impacts from the act of being labeled gifted a part from actually participating in gifted programming. The potential impact of being previously labeled gifted in elementary school is examined compared to those students who are in gifted programming throughout middle school.

3.2 Sample and Data

Data for this study comes from the North Carolina Department of Public Instruction for use in evaluations of the Disadvantaged Student Supplemental Fund. Student level data including standardized test scores and socio-demographic information are included in this research. A student cohort of sixth graders in 2004-2005 who had data from 2003-2004 through 2006-2007 is created for this analysis. Roster data from DPI from 2004-2005, 2005-2006, and 2006-2007 are matched to testing files from 2003-2004 through 2006-2007 to create this cohort. This data set is a significant improvement upon existing studies since individual student data for an entire state including specific classrooms and schools is available for analysis. The numerous controls available in this expansive data set enable a more detailed analysis of learning environments than in previous studies. The multiple time points permit a more advanced statistical method of hierarchical linear modeling growth curve analyses.

The dummy variable used to identify students who were gifted at any point in middle school (GiftedMS = 1) is the key variable of analysis. This is included at level two of the growth model since this variable does not change over time. The percentages of gifted students remain largely constant across the three years of middle school as displayed in Table 1.

Table 1: Percent of Gifted Middle School Students by Year

Year	Math	Reading
2004-2005	20.3%	20.2%
2005-2006	20.2%	20.3%
2006-2007	21.7%	21.9%

Although gifted policies have increased over time, most recently with the adoption of statewide Academically and Intellectually Gifted (AIG) Program Standards in July 2009 (N.C.G.S. § 115C-150.05-.08), during the years of analysis gifted programming varied considerably by each local school district. In a 2001 report prepared for the North Carolina Department of Public Instruction (DPI) and presented to the State Board of Education, schools were surveyed and DPI records documented to analyze access to advanced courses and gifted programs across North Carolina schools. This study was enacted in response to SL2000-67; “this legislation directed the State Board of Education to study underrepresentation of minority and at-risk students in Honors classes, Advanced Placement (AP) classes, and academically and intellectually gifted (AIG) programs; to evaluate whether a student is eligible for one of these classes or programs and how objective these criteria are and to explore the extent to which low academic expectations contribute to representation" (Darity , Castellino, Tyson, Cobb, & McMillen, 2001: i). They conclude that the proportional gap by race/ethnicity is "significant and widespread."

Findings show that most elementary and middle schools structure their AIG (gifted) program using a resource room (51%) and/or a heterogeneously-grouped classroom (49%). Other program structures include clustering and enrichment; the trend is that schools use different intensities and often multiple types of services. Screening instruments also vary. While most schools (51%) report that End of Grade Tests are used for the screening process, other cognitive abilities tests, teacher checklists, and the Otis Lennon School Ability Test are also used. At the elementary school level, an increase in the number of instruments used in the screening process is statistically significant to the number of students enrolled in the AIG program (Darity et al., 2001).

In addition to assessment data, teacher recommendations, grades, and student-self selection are also criteria for gifted identification. Table 2 below shows the distribution of reported identification of gifted students; clearly many schools use multiple methods of identification.

Table 2: Criteria Used for Identification and Placement

Criteria	Percent of Schools
Teacher Recommendation	90
EOG test scores	90
Cognitive/intelligence test	86
Grades	81
Self-selection (including parent request)	66
Student Work Portfolio	62
Standardized achievement test	53
Outside or independent assessment/evaluation (by parent request)	45
Other assessment procedure	36
Domain or skill-specific aptitude tests	13

From Increasing Opportunities to Learn via Access to Rigorous Courses and Program (Darity, Castellino, Tyson, Cobb & McMillen, 2001)

It is important to note that this analysis measures the impact of gifted student programming treatment, yet that treatment may not be uniform across local districts. There are no common standards for gifted students until the 2009-2010 school year leaving room for local interpretation of gifted programming. Previously the only requirement was for local education agencies to develop three year-plans for programming to be approved by the local school board and then sent to DPI for comment. The state has mandated, however, that identified students receive gifted programming since specific academically and intellectually gifted legislation was passed in 1996. Therefore, during the school years of this analysis there is likely variation as to what gifted students experience across districts and schools.

There are differences in the sample sizes of the math and reading analyses. This is due to the data received from NC DPI and students with adequate matches. Only those students with complete information (each characteristic) from 2003-2004 through 2006-2007 were accepted into the HLM model to run analyses; cases with missing data are dropped. Reading students had less eligible cases to include as seen in Table 3 below.

Table 3: Sample Sizes by Level

	Math	Reading
Level One	117,169	114,877
Level Two	40,832	38,391
Level Three	1,164	910

3.3 Measures

Dependent Variable

The dependent variable in both models is the student's end of grade test scores in reading or math. These standardized scores enable an examination of student academic achievement as measured by the statewide assessment required of all students in grades three through eight. Beginning with the 2005-2006 school year, the Math End of Grade tests were changed to match the 2003 North Carolina Mathematics Standard Course of Study which is the content standards for the state. Therefore, the 2004-2005 math data is rescaled to match the math End of Grade test scale changes that began in the 2005-2006 school year. The reading tests are not changed until the 2007-2008 school year to match the 2004 English and Language Arts Standards so no adjustment is necessary. All test data is trimmed according to DPI's concordance tables for each grade. These are tables that provide a range of reasonable scores for each End of Grade test by grade and year.

Individual Student Measures

The key independent variable is gifted student status in middle school, GiftedMS. There is also a dummy variable for students who had the gifted label (Gifted = 1) in the fifth grade of elementary school, GiftedG5.

A measure of previous student ability is needed to limit the bias in coefficients of individual student measures due to correlation with student ability (Hanushek, 1997). The previous fifth grade end of grade test scores in reading or math (2003-2004) are used to capture previous student achievement.

Socioeconomic status is often significant to differences in learning and a standard control variable in education studies. Proxies for socioeconomic status are parent education level and free and reduced lunch status. The education level of a child's parent, ranging from less than high school to college graduate, is included in the study. Free or reduced lunch eligibility is an indicator for income level. Free lunch students are those with family incomes at or below 130 percent of the poverty level, while reduced lunch students are those with family incomes between 130 and 185 percent of the poverty level ("National School Lunch Program," 2007). Data for both measures are taken from the DPI Student Activity Report.

Additional variables include student race/ethnicity, gender, limited English Proficient (current and previous), and age (underage or overage based on state cut-off birth date). Disabilities are also controlled using a dummy variable (Disability=1) if a student has cognitive, behavioral, sensory, high incidence, or severe disabilities.

Teacher Characteristics

The percentage of teachers with less than a bachelor's degree and with advanced degrees are included the propensity score match analyses. Teacher experience adds to effectiveness until a threshold of about twenty five years where effectiveness begins to decline (Klitgaard & Hall, 1974; Murnane & Phillips, 1981; Rosenholtz, 1986). Thus, teacher experience will be broken into segments including teachers with zero to three years experience, and those with four to ten years of experience. These variables are not included in the hierarchical linear models because growth models necessitate all time varying variables at level one. As previously explained, only non-time varying variables and school level means can be captured over time, thus teacher and classroom characteristics are not included in the growth models.

School Characteristics

There are potential differences across schools that could influence student achievement outcomes as well. Thus, percentages of race/ethnicity (Hanushek, Kain, & Rivkin, 2004) and lower income students (Fram, Miller-Cribbs, & VanHorn, 2007), school size based in average daily membership, and teacher turnover are included in the analysis. Level three IDs are also created from district and school codes to link this data to Levels One and Two.

Table 4 details the variables in the hierarchical linear model analysis.

Table 4: Variables in Hierarchical Linear Model Analysis Variables

Variable		Description
Level One	Reading Score or Math Score	Standardized End of Grade reading or math test score
	Time	Designates year from 2005-2007; Time = 0, -1, or -2
Level Two	Gifted MS	Student with the gifted label in middle school dummy GiftedMS=1 if Gifted
	Gifted G5	Student with gifted label in fifth grade dummy GiftedG5=1 if Gifted
	Male	Male = 1 if Male
	Black	Black = 1 if Black
	Other	Other = 1 if Hispanic, American Indian, Asian, or Multiracial
	FrLnch	FrLnch = 1 if Free Lunch dummy
	RedLnch	RedLnch = 1 if Reduced Lunch dummy
	FRLunchMiss	FRLunchMiss = 1 if Free or Reduced Lunch missing
	Ped_lesshs	Parent Education; Ped_lesshs = 1 if less than high school
	Ped_cols	Parent Education; Ped_cols = 1 if some college
	Ped_colg	Parent Education; Ped_colg = 1 if college graduate
	Pedmiss	Parent Education; Pedmiss = 1 if missing
	IsLep06	IsLep06 = 1 if currently Limited English Proficient
	WasLep06	WasLep06 = 1 if student was Limited English Proficient
	Exceptional	Exceptional = 1 if child has a disability (high incidence, cognitive, sensory, physical)
	Underage	Underage = 1 if student is one year or more younger than grade cut-off date
	Overage	Overage = 1 if student is one year or more older than grade cut-off date
	Zma_scoreG5 or Zrd_scoreG5	Standardized End of Grade reading or math test score from fifth grade (2003-2004)
Level Three	Black Students	Percentage of black students in the school
	Other Students	Percentage of other students in the school
	FRLunch Students	Percentage of students who qualify for free price lunch
	Redlnch	Percentage of students who qualify for reduced price lunch
	Turnover	Teacher Turnover
	Adm	School size using average daily membership

3.4 Analytic Approach

Since treatment of gifted students is not random, matched sampling is used to account for this potential bias in the analyses. Gifted and not-gifted elementary school students are matched in an effort to provide unbiased estimates of the treatment effects. Then two hierarchical linear models are used to quantify the potential effects on reading and math outcomes. Growth models estimate differences in student achievement of gifted and not-gifted students. Time is nested in students and students are nested in schools to determine the average treatment effect resulting from gifted programming and how this varies across race/ethnicity and income.

Propensity Score Matching

It is clear that randomized experiments are the gold standard of research; however since this is not feasible, propensity score analysis is used to estimate program effects (Rubin, 1973; Rosenbaum & Rubin, 1983; Rosenbaum & Rubin 1985). Thus, the treatment of gifted is matched to similar students based on observed characteristics. This selection technique has been shown to substantially remove bias from effect estimates (Cook, Shadish, & Wong, 2008). Matched sampling is a common technique used in education studies; however, PSM is not widely used in gifted student literature. In the most related study, Tong and Yewchuk (1996) match 39 gifted and talented students on gender and grade to examine sex-role orientation and self-concept. More recently, Sparfeldt (2007) surveys a matched sample of gifted and "average ability" students based on IQ scores to study vocational interests. In essence the technique identifies students for a comparison group that are similar on measured covariates to treated students, in this case gifted students.

In this research, two propensity score matches are conducted to address each subject area. Since gifted student assignment is not random and gifted students are more likely to be white and higher income, this sampling technique is used to provide unbiased estimation of the treatment effects. "The use of a study's sampling design to minimize initial differences between the control and comparison populations" (Cook, Shadish, & Wong, 2008: 745) is a critical aspect of this analysis. Selecting comparison students on their propensity to be labeled gifted can reduce bias in the estimates of the effects through the use of a more balanced sample. Propensity score analysis uses a logistic regression to create a value of the propensity for each student to receive the treatment, gifted in this analysis. A range of variables are used to create this single score or logit. The central idea is to match gifted students to students who had the propensity to be labeled gifted but were not so labeled. The propensity is measured through a composite of covariates that are combined through a logistic regression. The predicted values of the regression are the probability of a student being treated (Ravallion, 2001).

Previous test scores, race/ethnicity, gender, and income level are the individual level covariates in the matching of fifth grade students in 2003-2004. Teacher-related matching variables include percentages of teacher education level and years experience. Additionally, in order to account for differences in gifted programming across schools, total number of teachers, school size and percentages of student race/ethnicity and free reduced lunch are also included in the match. Since gifted education is a locally implemented policy, some variation is expected in identification and programming. Thus, students are matched according to individual, teacher, and school characteristics in the fifth grade; a sample of comparison students is selected for the statistical analysis.

A strength of matched samples is that assignment to treatment (gifted program) is independent of observed covariates (Ravallion, 2001). All the variables in the matching process can be used in the analysis and eliminate potential biases in the outcomes due to these characteristics. Students are then matched on the predicted probability of the logistic that provides the propensity to be labeled gifted. Nearest neighbor matching is commonly used to create the initial matched sample of students. "The observation in the nonparticipant sample that has the closest propensity score, measured by absolute differences in scores" is matched to a gifted student to enable more precise measures due to the proximity of scores (Ravallion, 2001: 139). This method is also compared to kernel and radius matching to determine the matching method resulting in the most balanced samples. Kernel matching can be conducted in two ways; with Gaussian or Epanechnikov kernels. A Gaussian kernel uses all non-treated cases compared to an Epanechnikov kernel which only uses non-treated cases within a fixed caliper of the treated unit (Sianesi, 2001). With both types, "all treated are matched with a weighted average of all controls with weights that are inversely proportional to the distance between the propensity scores of treated and controls (Baser 2006: 380-381)." Thus the weights of treated cases are equal to one, while weights of not-treated cases range above and below one depending on their propensity to have been treated. An advantage to kernel matching is that more control group cases are utilized in the match resulting in lower variance. Radius matching simply matches treated cases to not-treated cases within a certain, pre-determined neighborhood or radius of the propensity score of the treated case. This can result in numerous unmatched cases depending on the size of the radius and quality of the matches; yet this method can also result in quality matches due to a small radius (Becker & Ichino, 2002). Epanechnikov kernel matching is the method chosen for analysis as fully explained in the matched sample results in the following chapter.

Bearing in mind that variation in middle school outcomes, now based on these two equivalent groups, is the main thesis of this research, the extent to which students in gifted programs outperform similar students not in gifted programs is crucial. It is hypothesized that this difference is considerable as time progresses due to the additional resources and learning supports provided to those in the gifted program. Descriptive statistics of the matched middle school cohort are examined to determine the extent to which disparities exist in gifted participation and student performance. Hierarchical linear growth models are then analyzed to help elucidate whether or not there are significant gains in the performance outcomes of gifted students.

Hierarchical Linear Modeling

Multilevel models enable a more in-depth analysis than ordinary least squares regression (OLS) by distributing variance across levels. For instance, in a three level model of time, individual students, and schools, the variance is divided into separate levels instead of one larger number isolating time-variant, individual, and school effects. The clustering of variables enables an extension of traditional OLS. This improvement in standard error measurements increases the statistical power of the model (Gelman & Hill, 2007).

Seltzer (1995) uses Bernstein's slopes-as-outcomes multilevel analysis to examine equity of programs in education evaluation research. He argues that it is critically important to consider student's initial status and within-group slopes over time, not solely group averages. The use of hierarchical linear modeling allows the relationship between student predictors (race/ethnicity, income, etc.) and student outcomes to vary with organizational (school) contexts. Slopes vary with each student and are modeled as a result of different educational experiences. The multiple levels of HLM therefore enable comparisons of treatment effects within and across groups. “By employing an analytic approach that allows for the possibility that program effects may vary across sites, we are able to start moving beyond addressing questions concerning how beneficial a program is on average, to searches for those factors that appear to promote high levels of success…evaluators can begin to examine differences in the extent to which programs of interest eventuate in equitable distributions of achievement” (Seltzer, 1995: 303).

This is applicable to this research as student outcomes are expected to vary over time and by race/ethnicity and income. The different school environments in which learning takes place are controlled by the school level variables.

A cohort of students at three points in time with time varying characteristics, individual student-family factors, and school characteristics enable a deeper understanding of the gifted label on student outcomes. A growth curve model is used for analysis to examine student achievement over time. End of Grade test scores over time are the outcome variables with separate models for reading and math. Since time is level one, differences in rate of growth are at the first level and are uncorrelated with the observed, personal student characteristics (i.e. race/ethnicity, free lunch status) at level two. Covariates are seen in Table 5 by level.

Table 5: Hierarchical Linear Model Levels

***Level One:* Time Varying Characteristics**
***Level Two:* Individual Characteristics**
***Level Three:* School Characteristics**

After the cohort is built based on students in the sixth grade in 2004-2005 by matching student data files, variables are separated into the appropriate levels to conduct HLM analyses. Level One contains middle school students for each of the three years (denoted time) and their test scores for each year. Grade 8 is used as time = 0 (time = -1 for grade 7, and time = -2 for grade 6), thus the intercept is the difference in performance at the end of eighth grade. The time coefficient measures the effect on test score for each additional year.

Level Two is comprised of student demographic and family data. Student and family characteristics that do not change, such as race/ethnicity and free/reduced lunch status, are included at this level. The weights from the propensity score match (vweight) are also matched into the cohort data and included at level two. The Epanechnikov kernel match calculates weights for each observation; weighted averages of all students in the control group, gifted in fifth grade, are used to create the counterfactual outcome. "Weights depend on the distance between each individual from the control group and the participant observation for which the counterfactual is estimated (Caliendo and Kopeling, 2005: 11)." The weight increases as the distance decreases. Weighting observations is a technique commonly used to produce unbiased estimates of population parameters. As in this research, "suppose persons are selected with known probability [propensity score from the logistic regression] and then followed longitudinally over time…we have occasions at level 1 nested within persons at level 2. The only weight may be a level-2 weight, inversely proportional to the probability of selection of that person" (Scientific Software International, 2010). Finally school level characteristics, which are measured as the school means calculated from the three years of middle school, are included in Level Three.

A null model of each data set, math and reading, is run to determine the baseline from which to compare the more detailed models. The null model is also called the unconditional model. The intercept is estimated in a random coefficients model; the intercept varies at levels 2 and 3. In Table 6 below, 0.51 is the mean test score for student i in school j. As no controls are included in this estimate, the null model is used primarily to compare models and determine how much variance is explained as variables are added. There is no significant difference in the average reading achievement between gifted and not-gifted students in the null model. As gifted variables are added, this intercept becomes significant and the size of the coefficient increases. There is a significant difference in the predicted mean test scores of math students; these differences become negative as gifted variables are added to the null model.

Table 6: Preliminary Model Comparisons

		Math			Reading		
		Null Model	Basic 1	Basic 2	Null Model	Basic 1	Basic 2
Intercept	Intercept	.051*** (.012)	-0.224*** (0.011)	-0.238*** (0.011)	-0.005 (.0113)	-0.261*** (0.012)	-0.275*** (0.012)
	Gifted MS		1.297*** (0.011)	0.834*** (0.338)		1.185*** (0.012)	0.753*** (0.019)
	Gifted G5			0.560*** (0.018)			0.518*** (0.020)
Slope	Time Slope Intercept		-0.039*** (0.003)	-0.039*** (0.003)		-0.003* (.002)	-0.003* (0.002)

Robust standard error in parenthesis, *p<.05, **p<.01, ***p<.001

Although the gifted middle school and gifted grade five variables are highly correlated, 0.821 in math and 0.841 in reading, both variables are significant in each basic 2 model. The addition of the giftedG5 variable decreases the estimated effect of gifted middle school for both math and reading since there is likely an impact of both previous and current gifted programming on student outcomes. The inclusion of the gifted grade five variable enables an examination of the potentially different effect of previously gifted versus being currently gifted (gifted middle school) on student outcomes. Therefore the gifted in fifth grade dummy variable is included to control for students who were gifted in elementary school.

Two full math and reading models are then run with time, individual, and school level characteristics. The full model, divided into each level, is displayed below in Figure 2 (note Table 4 identifies each variable). Coefficients for key variables are displayed in Table 7 and the variance is shown in Tables 8 and 9.

Level-1 Model

Y = P0 + P1*(TIME) + E

Level-2 Model

P0 = B00 + B01*(BLACK) + B02*(OTHER) + B03*(FRLNCH) + B04*(REDLNCH) + B05*(FRLUNCHM) + B06*(PED_LHS) + B07*(PED_COLS) + B08*(PED_COLG) + B09*(PEDMISS) + B010*(MALE) + B011*(WASLEP) + B012*(ISLEP) + B013*(EX_DIS) + B014*(UNDERAGE) + B015*(OVERAGE) + B016*(EX_AIG05) + B017*(GIFTEDMS) + B018*(ZMA_SCOR) + R0

P1 = B10 + B11*(BLACK) + B12*(OTHER) + B13*(FRLNCH) + B14*(REDLNCH) + B15*(FRLUNCHM) + B16*(PED_LHS) + B17*(PED_COLS) + B18*(PED_COLG) + B19*(PEDMISS) + B110*(MALE) + B111*(WASLEP) + B112*(ISLEP) + B113*(EX_DIS) + B114*(UNDERAGE) + B115*(OVERAGE) + B116*(EX_AIG05) + B117*(GIFTEDMS) + B118*(ZMA_SCOR) + R1

Level-3 Model

B00 = G000 + G001(BLACK_3) + G002(OTHER_3) + G003(FRLNCH_3) + G004(REDLNCH) + G005(TURNOVER) + G006(ADM_3) + U00
B01 = G010
B02 = G020
…
B018 = G0180
B10 = G100 + G101(BLACK_3) + G102(OTHER_3) + G103(FRLNCH_3) + G104(REDLNCH) + G105(TURNOVER) + G106(ADM_3) + U10
B11 = G110
B12 = G120
…
B118 = G1180

Figure 2: Hierarchical Linear Model Equation by Level

The second full model, as shown above, adds the fifth grade test score. It is important to note that there are econometric problems that are addressed through the use of HLM. All time varying-characteristics must be at Level One in a growth model, thus individual characteristics at Level Two and school characteristics at Level Three do not change over time. As previously explained, the Level Two variables are from 2005-

2006, the middle year for students, and Level Three variables are an average of school level data across the three years. Error is portioned at the different levels to reveal explained variance across the different models.

All three variables are significant in the Full 2 model for both subjects except Gifted G5 in reading. The size of the gifted middle school coefficient decreases by about half when the fifth grade test score is included. Gifted G5 and the grade 5 reading test score are not highly correlated (.019) and Gifted G5 and the grade 5 math test score are moderately correlated (.499). Correlations between the fifth grade test score and gifted middle school variables are actually higher and also statistically significant. For math the correlation coefficient is 0.527 and for reading it is 0.534.

Table 7: Model Comparisons – Fixed Effect Math and Reading

		Math		**Reading**	
		Full 1	Full 2	Full 1	Full 2
Intercept	Intercept *B00*	0.696*** (0.021)	0.135*** (0.019)	0.513*** (0.018)	0.253*** (0.016)
	Gifted MS *B017*	0.388*** (0.019)	0.137*** (0.017)	0.353*** (0.020)	0.197*** (0.018)
	Gifted G5 *B016*	-0.024 (0.019)	0.100*** (0.015)	0.064** (0.020)	0.026 (0.018)
	Grade 5 Score *B018*		0.537*** (0.010)		0.464*** (0.009)
Slope	Time Slope Intercept *B10*	-0.073*** (0.008)	-0.062*** (0.008)	-0.037*** (0.009)	0.105*** (0.008)
	Gifted MS *B117*	0.008 (0.009)	0.012 (0.009)	0.002 (0.011)	0.081*** (0.009)
	Gifted G5 *B116*	-0.004 (0.009)	-0.010 (0.009)	0.007 (0.011)	0.015 (0.009)
	Grade 5 Score *B118*		-0.014** (0.005)		-0.227*** (0.004)

Robust standard error in parenthesis, *p<.05, **p<.01, ***p<.001

Intraclass Correlation Coefficients (ICC) and explained variances are used to compare increasingly more complex models. Variance coefficients for each of the five models are displayed in Tables 8 and 9; these are the same models as previously displayed in Tables 7.

Table 8: Model Comparisons – Variance Components Math

	Null Model	Basic 1	Basic 2	Full 1	Full 2
Intercept L1	0.727*** (0.085)	0.508*** (0.713)	0.496*** (0.704)	0.029*** (0.171)	0.131*** (0.362)
Explained L1 Variance		*0.302*	*0.318*	*0.960*	*0.986*
Intercept L1/L2	0.136*** (0.369)	0.091*** (0.302)	0.090*** (0.300)	0.015*** (0.123)	0.051*** (0.226)
Explained Variance L2		*0.332*	*0.338*	*0.889*	*0.925*

Standard Deviation in parenthesis

Table 9: Model Comparisons – Variance Components Reading

	Null Model	Basic 1	Basic 2	Full Model 1	Full Model 2
Intercept L1	0.701*** (0.837)	0.502*** (0.709)	0.491*** (0.701)	0.207** (0.455)	0.202*** (0.449)
Explained L1 Variance		*0.310*	*0.299*	*0.705*	*0.712*
Intercept L1/L2	0.128*** (0.358)	0.074*** (0.272)	0.073*** (0.270)	0.042*** (0.205)	0.024*** (0.155)
Explained Variance L2		*0.423*	*0.429*	*0.672*	*.811*

Standard Deviation in parenthesis

The ICC is 0.842 for the fully unconditional (null) math model and 0.845 for the reading null model. This is calculated from the covariance estimates of the null models shown above by dividing the Level One intercept by the sum of the initial and slope intercepts. Thus, the total variance between schools is about 84 percent of the total variance for each subject and the remaining 16 percent is at the student level. The proportion of explained variance is then calculated for the increasingly more complex models. Clearly, the addition of Level Two and Level Three characteristics explain increasing amounts of explained variance across students and between schools. The final math model that includes the math test score from grade five explains 99 and 93 percent of the variance across students and between schools; the explained variance for reading is lower, with just 71 and 81 percent of variance across students and between schools explained. The addition of the Level Two and Three control variables helps explain differences in achievement across students and between schools and therefore should be included in the final model (Full 2 Model). The one exception is Gifted 05 in the reading Basic 2 model. There is a slight decrease in explained variance of about 1 percent for students; however, this is the only portion of the models where the variation decreased as variables were added.

Based on the previous considerations, the second full models in Table 7 are used for full analysis. This model includes the most characteristics: student dummies for gifted in middle school, gifted in fifth grade, and students' standardized test scores from fifth grade are all used. Time is nested in students who are in turn nested in schools. The dependent variable, Y, is the standardized math or reading test score. The years of school data are run using test scores for each point in time. Random slopes are used in the model to capture changes in growth over time in test scores in addition to the fifth grade test scores. Since TIME = 0 for eighth grade, the intercept measures the effect on the test score at the end of eighth grade, the final year of middle school. The stochastic part of the model is represented by E, R, and U, for time, individual students, and schools levels, respectively. Robust standard errors are calculated. Findings are detailed in the next chapter.

CHAPTER FOUR: RESULTS AND FINDINGS

This chapter details results of the statistical analyses used in this study to determine the extent to which gifted programming impacts student outcomes. First, the comparison groups from the propensity score matches are presented including the matching technique of choice. Descriptive data of matched cohorts of North Carolina students are examined to compare gifted and not-gifted students by race/ethnicity and income. Then the hierarchical linear growth model results from these analyses are detailed. Finally the findings are discussed and considerations for the unexpected results explored.

4.1 Matched Sample: Comparison Group

Gifted and not-gifted elementary school students are matched in an effort to provide unbiased estimates of the treatment effects. Since students are not randomly assigned to gifted programs, matching techniques are needed to create two equivalent groups to estimate unbiased effects of being in a gifted program. Four types of propensity score matches are run to determine the most balanced sample: nearest neighbor without replacement, Gaussian kernel matching, Epanechnikov kernel matching, and radius matching.

Epanechnikov kernel matching is the matching approach used for the final growth models as explained in more detail below. In order to complete the matches data from 2003-2004, the students' fifth grade year of elementary school, are pulled out from the cohort data set of 2003-2004 to 2006-2007. Comparisons of the original samples used for math and reading matches (2003-2004 data) are displayed in Table 10 below. There are large numbers of not-treated cases which helps facilitate high-quality matches (Ravallion, 2001).

Table 10: 2003-2004 Data Used for Propensity Score Matches

	Reading		**Math**	
	N	%	N	%
Gifted	8,141	19.07	9,372	18.04
Not-Gifted	34,556	80.93	42,577	81.96
Total	42,697	100	51,949	100

Areas of common support are present in both reading and math samples. The distribution of test scores for treated and not-treated cases is graphed in Figures 3 and 4 for each subject.

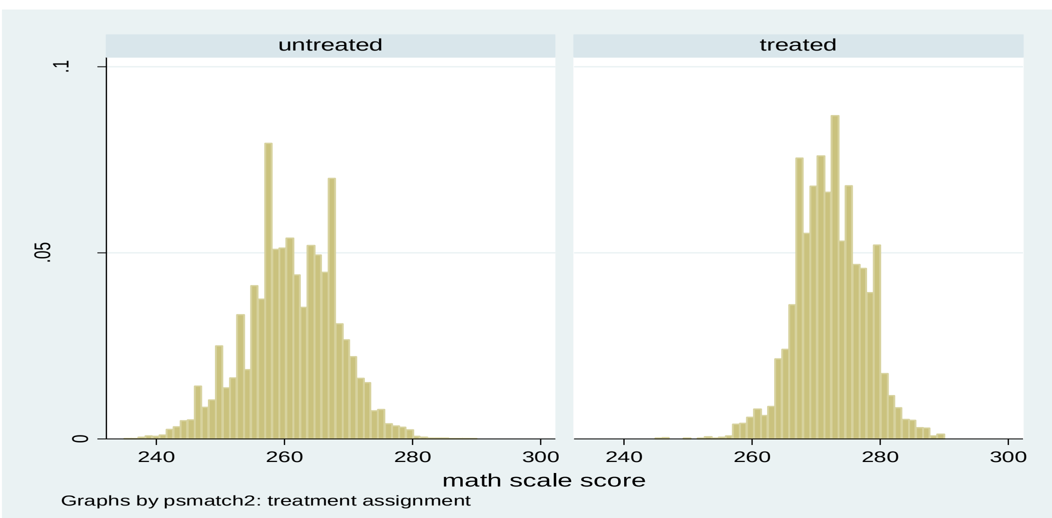


Figure 3: Math Common Support

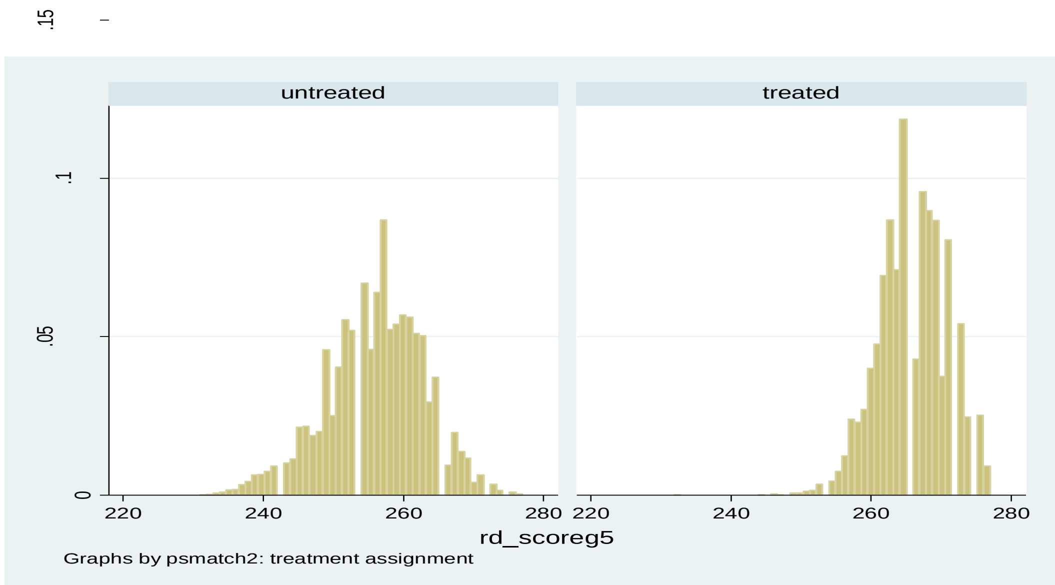


Figure 4: Reading Common Support

Since there are areas of overlap for both math and reading, adequate matches are expected from the propensity score matching techniques.

Logistic regressions are run in Stata to calculate propensity scores for fifth grade students. The outcome variable is the treatment variable – a gifted dummy variable (equal to one if in a gifted program) for 2003-2004, i.e., the fifth grade of elementary school. Covariates are the reading or math fifth grade score, race/ethnicity, free and reduced lunch status, and gender. Additional variables are also included to control for potential differences across schools; these include school averages of teacher education level, teacher years of experience, total number of teachers in the school, school size, and school percentages of students in poverty and in race/ethnicity groups. The complete logistic regression is displayed below.

gifted grade 5 = test score grade 5, black, other, freelunch, redlunch, male, percent below bachelors, percent advanced degree, percent teachers 0-3 years experience, percent teachers 4-10 years experience, total teachers 2004, average daily membership 2004, percent free or reduced lunch mean 2004, percent black mean 2004, percent other mean 2004
Table 4 explains each variable used in the analysis.

Figure 5: Logistic Regression

The estimated propensity score for each student is predicted from the logistic regression (see Appendix A). The propensity scores are used to run the psmatch2 command in Stata. Psmatch2 runs multiple types of Mahalanobis and propensity score matching techniques that approximate standard errors on treatment effects. The first technique utilized is nearest neighbor matching without replacement. The order of cases is randomized in Stata so the estimates are not affected by the order in which cases are matched. A new sub-sample is created based on matching each gifted student to one not-gifted student in the sample. This reduces the sample size considerably since not-gifted students without a match are dropped from the analysis. Kernel matching, both Gaussian and Epanechnikov, results in more cases because students are assigned weights based on the kernel-weighted average of the outcome of all non-treated cases. Radius matching also results in larger sample sizes than nearest neighbor matching without replacement since matches are based on a range of propensity scores.

Reduction biases are calculated for each of the four matching techniques for both math and reading. The covariates of gifted students in the original data file and not-gifted students in the matched file are compared to determine the percent of bias reduction for each match. Bias reductions are calculated as the reduction in the difference in means of the variables of the treated vs. non-treated cases compared to the difference in means of the treated vs. matched samples as seen in Figure 6.

$$\frac{|\text{Not-Treated Students} - \text{Treated Students}|}{|\text{Matched, Not-Treated Students} - \text{Treated Students}|}$$

Figure 6: Bias Reduction Calculation

Epanechnikov kernel matching results in the highest bias reductions; results are listed in Table 11 below. Nearest neighbor matching has the lowest percent reductions in bias of the three models. This is consistent with the findings of Henry and Yi (2009) in a study that assessed fourteen propensity score matching designs.

Table 11: Epanechnikov Kernel Matching Bias Reductions - Math

Variables	**Absolute Value of Differences in Means: All Not-Gifted Students and Gifted Students**	**Absolute Value of Differences in Means: Matched Students and Gifted Students**	**Reduction in Bias**
Math score	12.4838	0.2141	98.28%
Gifted (Treatment Variable)	1	1	0.00%
White	0.2561894	0.0147424	94.25%
Black	0.2264544	0.011306	95.01%
Other	0.0297349	0.0034363	88.44%
Male	0.0146791	0.0108888	25.82%
free lunch	0.2953433	0.0133336	95.49%
reduced lunch	0.0546519	0.0012456	97.72%
full pay	0.3499952	0.0145792	95.83%
parent less than high school	0.0873358	0.0139201	84.06%
parent high school only	0.2921616	0.0804422	72.47%
parent some college	0.0487126	0.0285645	41.36%
parent college grad	0.4290735	0.1237005	71.17%
parent edu missing	0.0008636	0.0007737	10.41%
tchrsstayed04	1.9735	0.02725	98.62%
tchr retention 04	0.81205	0.0532	93.45%
per nbpts 04	1.264037	0.153973	87.82%
per below bachelors	0.0015083	0.0001593	89.44%
per advanced degrees	1.35492	0.05891	95.65%
percent 0-3 years experience	0.87723	0.0489	94.43%
percent 4-10 years experience	1.12841	0.08317	92.63%
percent 11+ years exp	1.78404	0.64799	63.68%
total teachers 04	2.14521	0.15138	92.94%
school size	38.8404	0.1911	99.51%
percent free/reduced lunch	11.65955	0.21429	98.16%
black mean	4.7513	0.12971	97.27%
white mean	5.03509	0.3014	94.01%
other mean	0.28379	0.17169	39.50%
Propensity score	0.4343276	0.0046764	98.92%

Reading results are very similar; Epanechnikov kernel matching also produces the largest reductions in bias. The full reading bias reduction table is included in Appendix B. As a result of the propensity score matches, there are now comparable treated and not treated cases to be analyzed in the hierarchical linear models. Tables 12 and 13 below display descriptives of the reading and math Epanechnikov kernel matched samples for 2003-2004 data.

Table 12: Math Descriptives – Epanechnikov Matched Sample

	N	Minimum	Maximum	Mean	Std. Deviation
math scale score	51949	235	291	263.50	8.570
ex_aig2	51949	0	1	.18	.385
White	51949	0	1	.60	.489
Black	51949	0	1	.29	.452
Other	51949	0	1	.11	.316
Frlnch	51949	0	1	.35	.476
Redlnch	51949	0	1	.09	.284
Fullpay	51949	0	1	.57	.496
ped_lhs	51949	0	1	.09	.279
ped_hson	51949	0	1	.39	.489
ped_cols	51949	0	1	.23	.419
ped_colg	51949	0	1	.29	.454
Male	51949	0	1	.49	.500
Pr(ex_aig05)	50988	.00	1.00	.1818	.25498
psmatch2: Treatment assignment	50988	0	1	.18	.386
psmatch2: Common support	50988	1	1	1.00	.000
psmatch2: weight of matched controls	50988	.01	11.12	.3635	.71449

Table 4 explains each variable used in the analysis.

Table 13: Reading Descriptives – Epanechnikov Matched Sample

	N	Minimum	Maximum	Mean	Std. Deviation
read scale score	42321	231	277	257.81	7.717
ex_aig2	42697	0	1	.19	.393
White	42697	0	1	.60	.490
Black	42697	0	1	.28	.451
Other	42697	0	1	.12	.323
Frlnch	42697	0	1	.34	.474
Redlnch	42697	0	1	.08	.273
Fullpay	42697	0	1	.58	.494
ped_lhs	42697	0	1	.09	.282
ped_hson	42697	0	1	.38	.486
ped_cols	42697	0	1	.22	.413
ped_colg	42697	0	1	.31	.463
Male	42697	0	1	.49	.500
Pr(ex_aig05)	41917	.00	.97	.1919	.23197
psmatch2: Treatment assignment	41917	0	1	.19	.394
psmatch2: Common support	41917	1	1	1.00	.000
psmatch2: weight of matched controls	41917	.01	6.38	.3837	.57961

Table 4 explains each variable used in the analysis.

As previously explained kernel matching and radius matching keeps many more cases than nearest neighbor matching; however, there are some students who fall outside of the region of overlapping test scores for gifted and not-gifted fifth grade students. There are 961 math cases and 780 reading cases dropped in these matches because the math and reading test scores do not fall in the area of common support for treatment. The descriptive statistics of the students who are not matched are shown below in Tables 14 and 15.

Table 14: Math Descriptives – Epanechnikov Unmatched Sample

	N	Minimum	Maximum	Mean	Std. Deviation
math scale score	961	238	286	262.10	8.275
ex_aig2	961	0	1	.11	.311
white	961	0	1	.65	.479
black	961	0	1	.29	.454
other	961	0	1	.06	.246
frlnch	961	0	1	.35	.477
redlnch	961	0	1	.11	.313
fullpay	961	0	1	.54	.499
ped_lhs	961	0	1	.14	.344
ped_hson	961	0	1	.43	.495
ped_cols	961	0	1	.22	.418
ped_colg	961	0	1	.21	.408
male	961	0	1	.50	.500
Pr(ex_aig05)	0				
psmatch2: Treatment assignment	0				
psmatch2: Common support	961	0	0	.00	.000
psmatch2: weight of matched controls	0				

Table 4 explains each variable used in the analysis.

Table 15: Reading Descriptives – Epanechnikov Unmatched Sample

	N	Minimum	Maximum	Mean	Std. Deviation
read scale score	777	234	277	256.50	7.987
ex_aig2	780	0	1	.13	.333
White	780	0	1	.65	.478
Black	780	0	1	.29	.454
Other	780	0	1	.06	.243
Frlnch	780	0	1	.35	.478
Redlnch	780	0	1	.10	.302
Fullpay	780	0	1	.55	.498
ped_lhs	780	0	1	.14	.348
ped_hson	780	0	1	.39	.488
ped_cols	780	0	1	.24	.430
ped_colg	780	0	1	.22	.417
Male	780	0	1	.51	.500
Pr(ex_aig05)	0				
psmatch2: Treatment assignment	0				
psmatch2: Common support	780	0	0	.00	.000
psmatch2: weight of matched controls	0				
Valid N (listwise)	0				

Table 4 explains each variable used in the analysis.

4.2 Descriptive Statistics: Differences in Student Outcomes

The literature suggests that there are significant disparities in students identified for gifted programming; however, existing studies provide little information on the characteristics of students in gifted programs (Delcourt et al., 2007) nor examine the potential academic gaps within the gifted student population (Ford et al., 2008). Therefore, descriptive information of gifted students participating in gifted programs and those who do not are compared in grades six, seven, and eight to offer empirical evidence and quantify these expected differences.

Differences in students by race/ethnicity and income are explored to reveal the extent to which disparities exist in this student cohort from statewide data. Based on previous work (Delcourt et al., 2007; VanTassel-Baska, Feng, & Evans, 2007), there is expected to be more white and higher income students in gifted programming; these student populations are quantified in Table 16 below. Gifted students are 16 percent of the student population. The gifted and not-gifted columns break down percentages of gifted students by race/ethnicity and income.

Table 16: Description of Middle School Students - Math

Student Group	**Gifted**	**Not-Gifted**
All Students	16%	84%
Black Students (31%)	12%	35%
White Students (58%)	79%	54%
Lower Income Students (42%)	16%	47%
Higher Income Students (55%)	79%	48%

Although black students comprise over 30 percent of the student population, only 12 percent of gifted students are black while 79 percent of gifted students are white even though white students are below 60 percent of the population. Similarly, more than 40 percent of students are lower income, however only 16 percent of these students are gifted. Higher income students are about 55 percent of the student population and 79 percent of gifted students are higher income. This data supports previous research that black and lower income students are less likely to be labeled gifted.

Figure 7 below displays breakdowns of race/ethnicity and income for gifted and not-gifted students. Each set of columns totals the percent of black, white, low income and higher income students in the middle school population by gifted status.

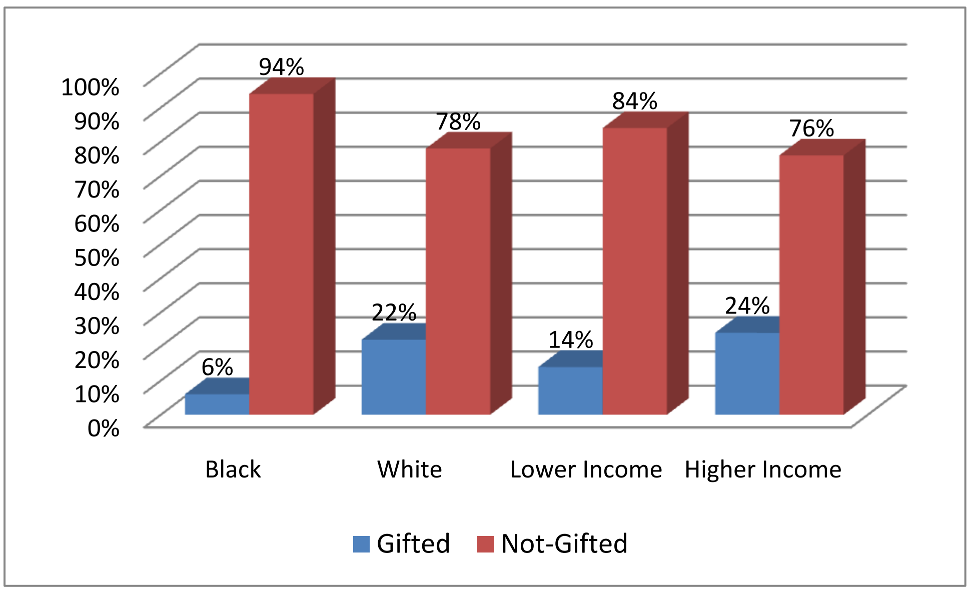


Figure 7: Gifted Status within Race/Ethnicity and Income - Math

When looking at black students, only 6% are gifted while 94% are not. This contrasts the white student population with 22% of students who are gifted and 78% who are not. Similar trends, exist when examining gifted students by income level. About 14% of lower income students are gifted compared to 24% of higher income students receiving programming.

Similar comparisons are found when examining gifted student populations by race/ethnicity and income for the reading outcome data as seen in Table 17.

Table 17: Description of Middle School Students - Reading

Student Group	Gifted	Not-Gifted
All Students	21%	79%
Black Students (28%)	10%	33%
White Students (60%)	81%	54%
Lower Income Students (39%)	14%	44%
Higher Income Students (55%)	80%	49%

Percentages are very similar to those of math middle school students. The percentages of black and lower income gifted students are low compared to their white and higher income classmates; gifted programming is not representative of student populations.

Breakdowns of race/ethnicity and income within reading gifted and not-gifted students are quite similar to math as seen in Figure 8 below.

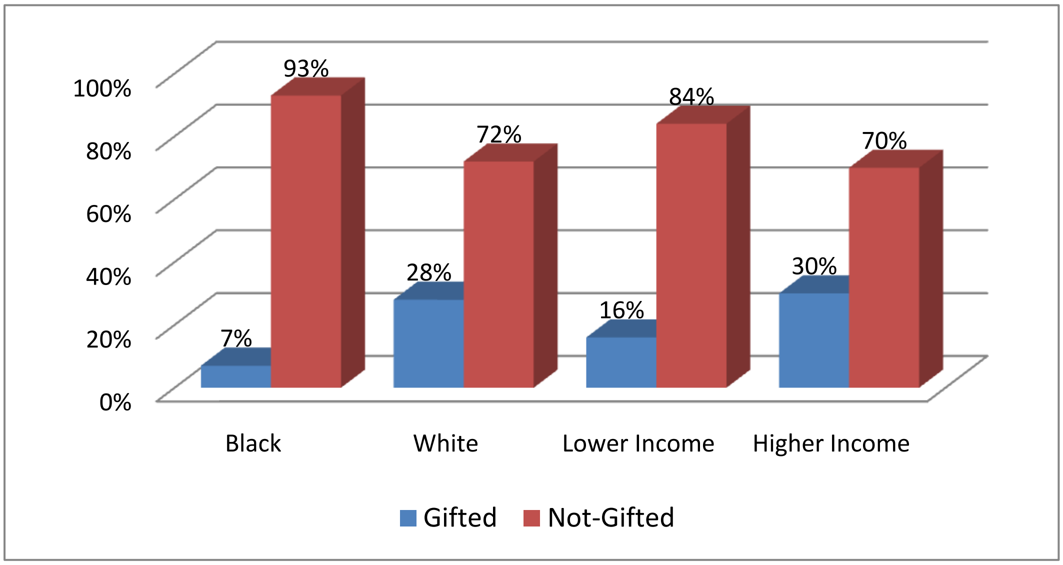


Figure 8: Gifted Status within Race/Ethnicity and Income - Reading

While there is a one percent increase in the number of black students who are gifted, there is a six percent increase in the proportion of white students who are gifted. The percentage of lower income students increases by two percent while the percentage of students who are higher income increases by six.

Tables 18 and 19 provided mean characteristics of gifted and not-gifted students by race/ethnicity and income level. Descriptive values for variables included in the HLM are below.

Table 18: Student Characteristics by Race/Ethnicity – Math

	Gifted		**Not-Gifted**	
	White	Black	White	Black
Middle School (MS) Math Scale Test Score	369.87 (4.91)	366.98 (5.87)	359.81 (8.05)	354.18 (7.85)
MS Math Standardized Test Score	1.16 (.54)	0.83 (.64)	0.03 (.88)	-0.60 (.85)
Free Lunch	5.6 (.23)	32.9 (.47)	20.8 (.41)	57.8 (.49)
Reduced Lunch	3.4 (.18)	10.5 (.31)	7.7 (.27)	11.0 (.31)
Parent Less than High School	0.7 (.08)	1.0 (.10)	5.4 (.42)	6.1 (.24)
Parent Some College	16.8 (.37)	24.7 (.43)	22.7 (.42)	21.2 (.41)
Parent College Graduate	60.5 (.49)	44.1 (.50)	27.7 (.45)	15.7 (.36)
Disabilities	0.2 (.05)	0.3 (.05)	12.3 (.33)	11.9 (.32)
Male	46.6 (.50)	41.0 (.49)	48.8 (.50)	46.4 (.50)
Is Limited English Proficient	0.0 (.01)	0.0 (.00)	0.1 (.03)	0.2 (.05)
Was Limited English Proficient	.01 (.02)	0.1 (.04)	0.2 (.04)	0.1 (.04)
Underage	1.2 (.11)	2.4 (.153)	0.4 (.07)	1.1 (.10)
Overage	7.7 (.27)	2.5 (.157)	22.4 (.42)	28.5 (.45)
Grade 5 Math Test Score	272.81 (5.04)	269.91 (5.20)	262.89 (6.68)	258.0 (6.72)
Grade 5 Math Standardized Test Score	1.25 (.63)	0.889 (.65)	0.02 (.83)	-0.59 (.84)
N = 117,169 for time (Level 1) and N = 40,832 for students (Level 2)				

Means are listed with standard deviations in parentheses.

Table 19: Student Characteristics by Race/Ethnicity - Reading

	Gifted		Not-Gifted	
	White	Black	White	Black
MS Reading Scale Test Score	271.24 (5.48)	267.90 (5.64)	262.43 (7.19)	257.42 (7.42)
MS Reading Standardized Test Score	1.05 (.62)	0.63 (.65)	-0.15 (.84)	-0.63 (.88)
Free Lunch	0.5 (.21)	29.6 (.46)	20.1 (.40)	56.2 (.50)
Reduced Lunch	0.03 (.16)	9.2 (.29)	7.1 (.26)	11.0 (.31)
Parent Less than High School	0.70 (.08)	1.0 (.10)	5.3 (.23)	6.5 (.25)
Parent Some College	15.2 (.36)	22.8 (.42)	22.1 (.42)	20.5 (.40)
Parent College Graduate	63.6 (.48)	47.1 (.50)	28.9 (.45)	16.5 (.37)
Disabilities	0.2 (.48)	0.2 (.05)	12.8 (.33)	13.2 (.34)
Male	48.1 (.50)	42.0 (.49)	50.0 (.50)	47.7 (.50)
Is Limited English Proficient	0.0 (.01)	0.0 (.00)	0.1 (.03)	0.3 (.06)
Was Limited English Proficient	.01 (.03)	0.30 (.06)	0.2 (.05)	0.2 (.04)
Underage	1.3 (.11)	2.9 (.17)	0.4 (.07)	1.2 (.11)
Overage	8.0 (.27)	2.8 (.17)	21.5 (.41)	28.3 (.45)
Grade 5 Reading Test Score	266.25 (4.73)	263.47 (4.98)	262.89 (6.33)	253.15 (6.57)
Grade 5 Reading Standardized Test Score	1.09 (.62)	0.73 (.65)	0.02 (.83)	-0.62 (.86)
N = 114,877 for time (Level 1) and N = 38,891 for students (Level 2)				

Means are listed with standard deviations in parentheses.

Tables 20 and 21 show similar comparisons of gifted and not-gifted by income. Descriptive information for these math and reading characteristics are below.

Table 20: Student Characteristics by Income – Math

	Gifted		**Not-Gifted**	
	Lower Income	Not Lower Income	Lower Income	Not Lower Income
MS Math Scale Test Score	367.78 (5.49)	369.67 (5.05)	355.10 (8.10)	359.92 (7.91)
MS Math Standardized Test Score	0.92 (.60)	1.15 (.56)	-0.49 (.88)	0.05 (.87)
Black	35.0 (.48)	7.9 (.27)	50.9 (.50)	18.1 (.39)
White	46.8 (.49)	85.6 (.35)	32.8 (.47)	75.2 (.43)
Parent Less than High School	4.8 (.21)	0.4 (.07)	12.0 (.33)	2.9 (.17)
Parent Some College	29.3 (.46)	16.1 (.37)	19.5 (.40)	23.3 (.42)
Parent College Graduate	24.9 (.43)	62.4 (.49)	9.6 (.30)	32.5 (.47)
Disabilities	0.3 (.06)	0.2 (.05)	12.6 (.33)	10.6 (.31)
Male	43.4 (.50)	45.9 (.50)	46.8 (.50)	49.0 (.50)
Is Limited English Proficient	0.5 (.07)	0.1 (.02)	4.4 (.20)	0.8 (.09)
Was Limited English Proficient	1.6 (.13)	0.2 (.04)	2.0 (.14)	0.5 (.07)
Underage	1.7 (.13)	1.5 (.12)	0.8 (.09)	0.9 (.09)
Overage	7.4 (.26)	7.1 (.26)	32.7 (.47)	18.1 (.39)
Grade 5 Math Test Score	270.69 (5.03)	272.76 (5.16)	258.85 (6.90)	262.88 (6.71)
Grade 5 Math Standardized Test Score	0.99 (.63)	1.24 (.64)	-0.49 (.86)	0.01 (.83)
N = 117,169 for time (Level 1) and N = 40,832 for students (Level 2)				

Means are listed with standard deviations in parentheses. Low income is students eligible for free or reduced lunch meals. Not low income is students who are not eligible.

Table 21: Student Characteristics by Income – Reading

	Gifted		**Not-Gifted**	
	Low Income	Not Lower Income	Low Income	Not Lower Income
MS Reading Scale Test Score	268.48 (5.6)	271.41 (5.54)	258.12 (7.60)	262.80 (7.06)
MS Reading Standardized Test Score	0.69 (.64)	1.04 (.63)	-0.55 (.90)	0.00 (.83)
Black	31.9 (.47)	6.9 (.25)	49.4 (.50)	17.5 (.38)
White	48.0 (.50)	85.7 (.35)	32.5 (.47)	75.4 (.43)
Parent Less than High School	5.6 (.23)	0.4 (.07)	12.9 (.34)	2.8 (.17)
Parent Some College	27.9 (.45)	14.6 (.35)	18.8 (.39)	22.6 (.42)
Parent College Graduate	26.4 (.44)	64.8 (.48)	9.6 (.30)	33.6 (.47)
Disabilities	0.2 (.05)	0.2 (.05)	13.4 (.34)	11.0 (.31)
Male	44.8 (.50)	47.3 (.50)	48.0 (.50)	50.1 (.50)
Is Limited English Proficient	0.7 (.08)	0.1 (.02)	5.3 (.22)	0.9 (.10)
Was Limited English Proficient	2.3 (.11)	0.3 (.06)	2.3 (.15)	0.6 (.08)
Underage	1.3 (.22)	1.9 (.14)	0.8 (.09)	0.9 (.09)
Overage	7.3 (.26)	7.5 (.26)	32.6 (.47)	17.2 (.38)
Grade 5 Reading Test Score	263.67 (5.07)	266.15 (4.74)	253.63 (6.69)	257.76 (6.29)
Grade 5 Reading Standardized Test Score	0.75 (.66)	1.08 (.62)	-0.56 (.87)	-0.02 (.82)
N = 114,877 for time (Level 1) and N = 38,891 for students (Level 2)				

Means are listed with standard deviations in parentheses. Students eligible for free or reduced lunch meals are lower income students.

Finally, school characteristics are overviewed for reading and math data sets in Tables 22 and 23.

Table 22: School Characteristics – Math

Percent Black Students	23.82 (19.46)
Percent White Students	61.27 (22.46)
Percent Free Lunch Students	19.56 (10.96)
Percent Reduced Lunch Students	6.10 (3.46)
School Size	660.97 (242.15)
Teacher Turnover	20.66 (6.80)
N = 1,164 (Level 3)	

Means are listed with standard deviations in parentheses.

Table 23: School Characteristics – Reading

Percent Black Students	24.47 (19.22)
Percent White Students	60.71 (22.35)
Percent Free Lunch Students	19.02 (10.72)
Percent Reduced Lunch Students	6.30 (3.99)
School Size	716.38 (251.35)
Teacher Turnover	21.32 (6.98)
N = 910 (Level 3)	

Means are listed with standard deviations in parentheses.

Comparisons over time are also significant to the analysis. When examining test scores over time, gifted and not-gifted students increase at about the same rate each year as evident in Figure 9.

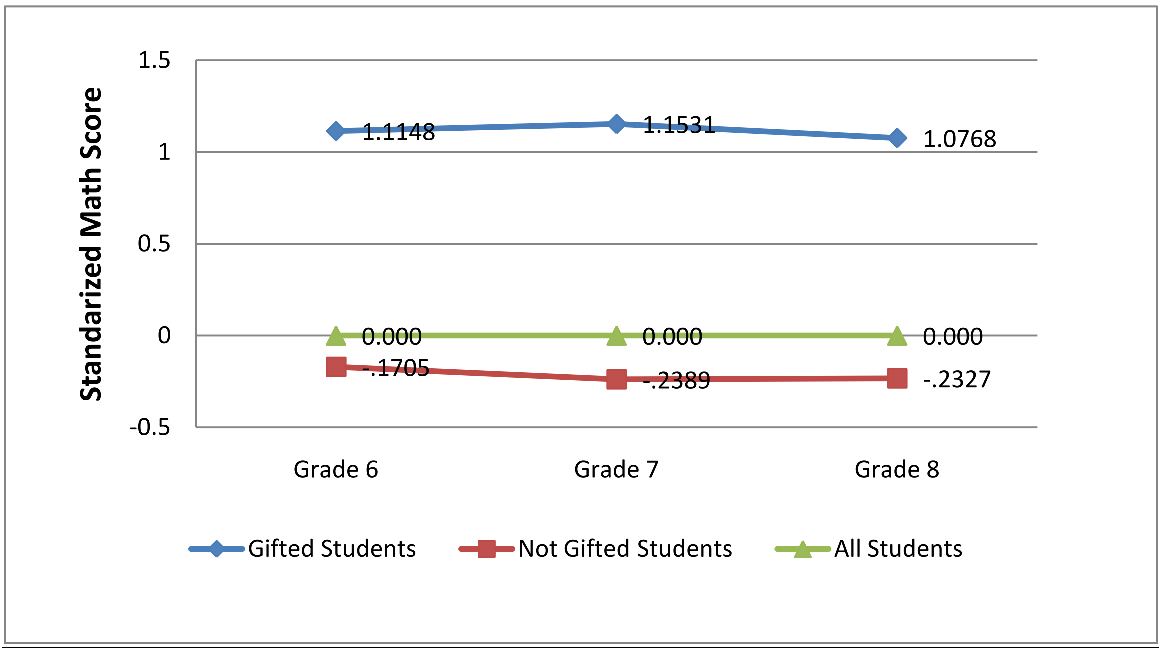


Figure 9: Test Performance by Year - Math

Gifted students have higher initial test scores than not-gifted students, yet contrary to expectations, they actually do not increase in performance at much different rates than not-gifted students. This suggests that although students who are gifted in elementary school begin with higher test scores than their not-gifted peers, the middle school learning environment does not result in major increases in their rate of growth, as measured by End of Grade tests.

Figure 10 shows that reading comparisons of student performance over time are even flatter for reading students.

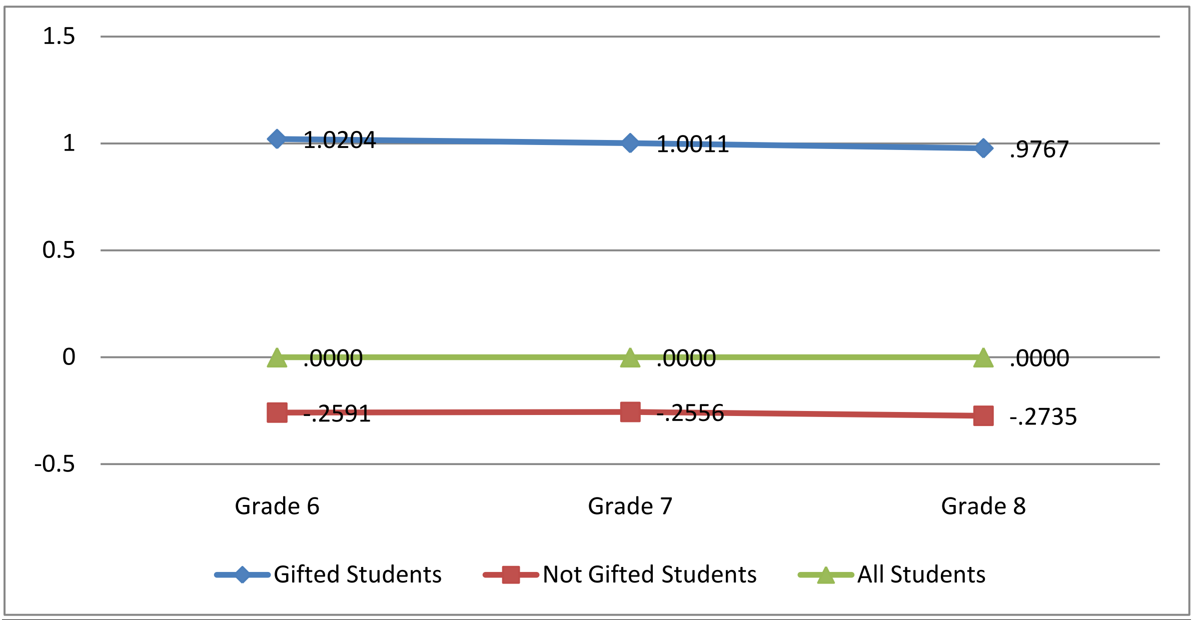


Figure 10: Performance by Year - Reading

Although gifted students have higher test scores than their not-gifted peers, there is not a larger increase over time in performance by gifted students.

Examining student performance by race/ethnicity is a central component of this research. Since black students comprise a disproportionally low percentage of the gifted population, black gifted students are compared to other gifted students in Figures 11 and 12 over middle school grades. Black not-gifted students are also plotted to compare their performance to their not-gifted peers.

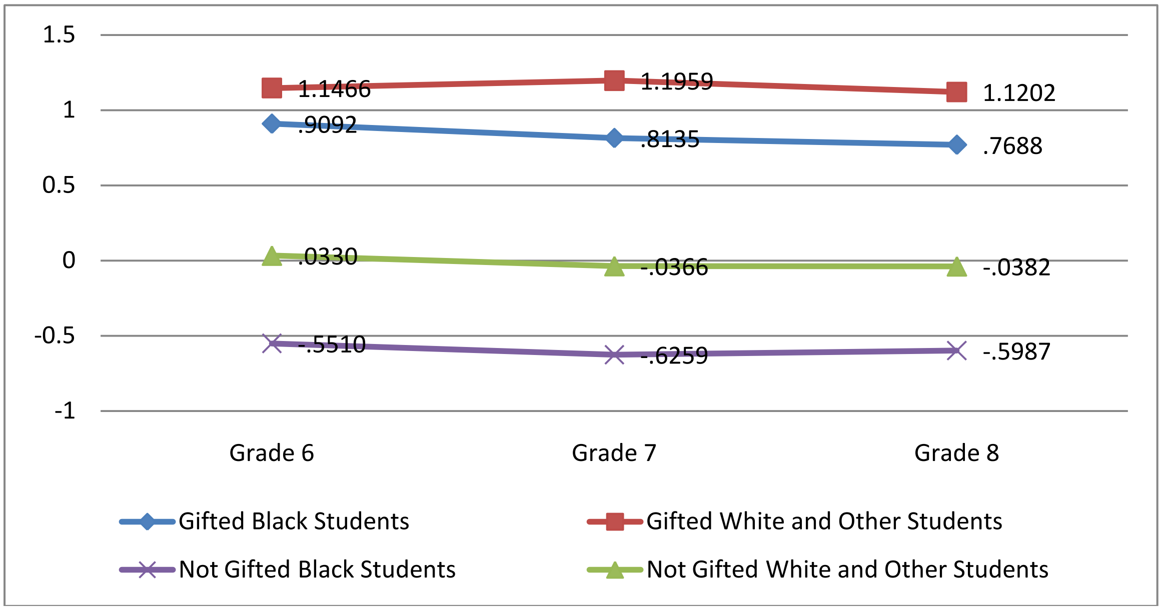


Figure 11: Performance by Race/Ethnicity - Math

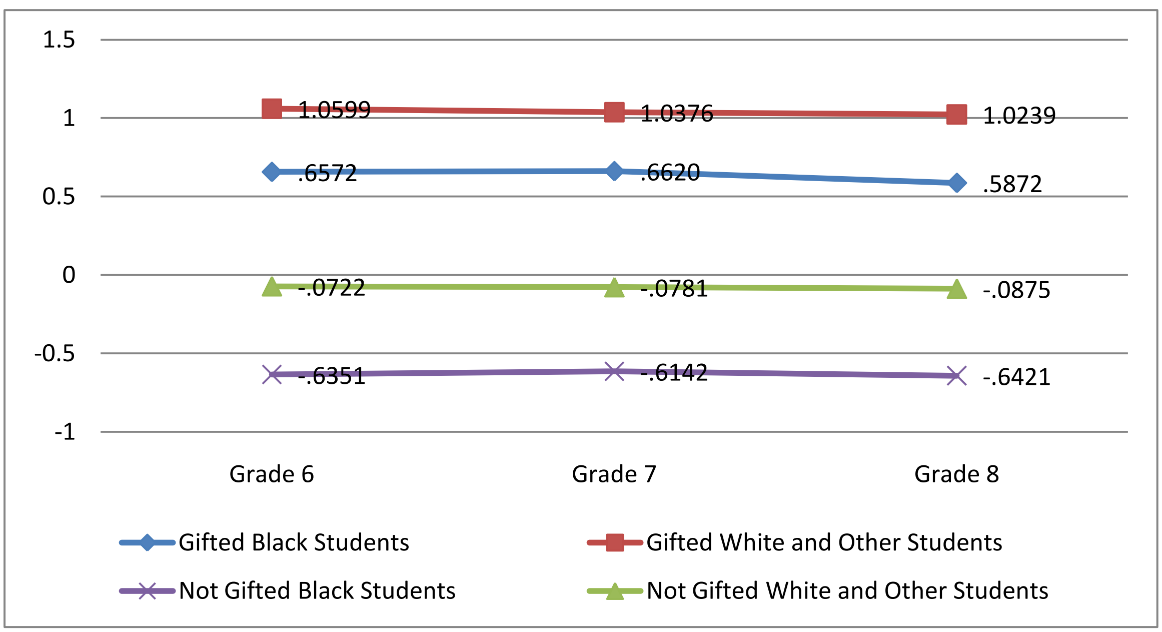


Figure 12: Performance by Race/Ethnicity - Reading

Changes in achievement by race/ethnicity are also flat over time. Student outcomes are stratified by gifted and status and race, yet there is a clear gap between the performance of gifted and not gifted students in the data. Finally, the test scores of students over time are graphed by income in Figures 13 and 14.

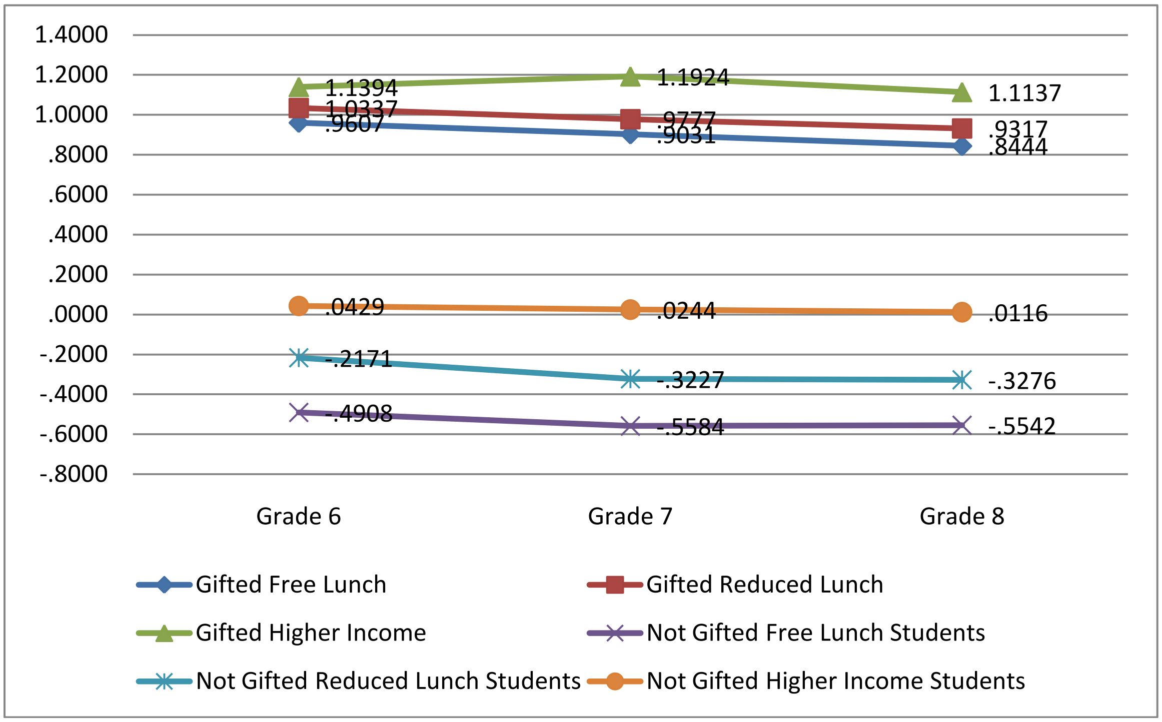


Figure 13: Performance by Income - Math

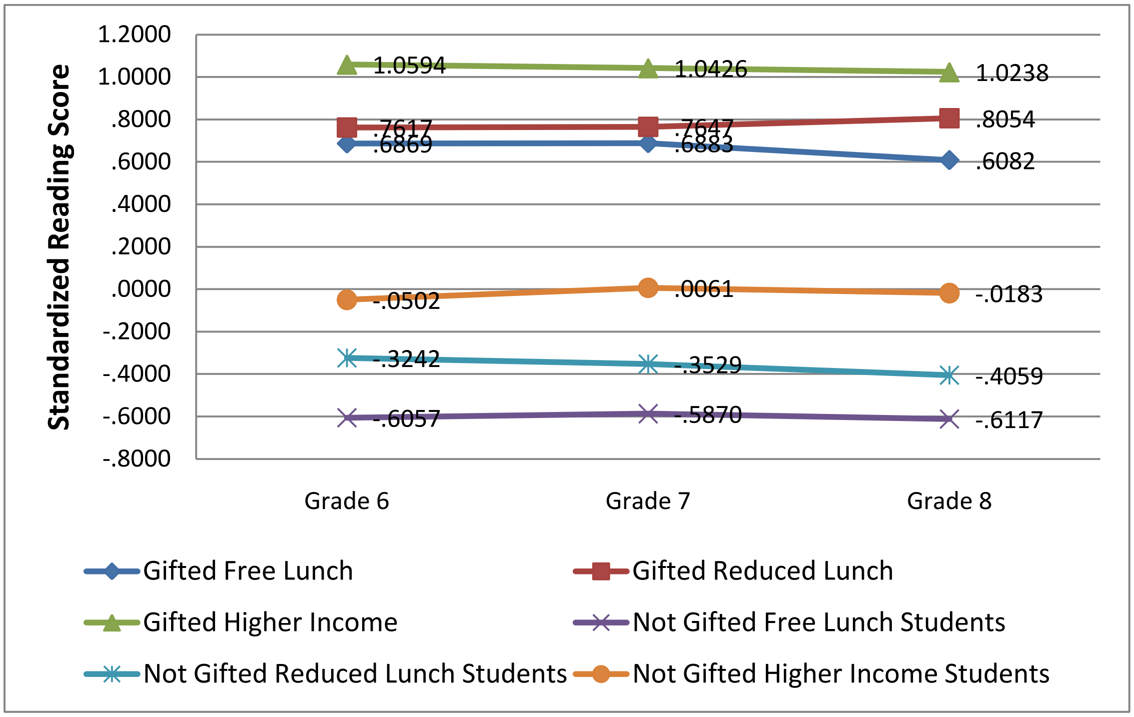


Figure 14: Performance by Income - Reading

These results are similar to graphs of gifted students and gifted students by race/ethnicity. There is not a large change in student performance over time by free, reduced, or full pay lunch status. In the cohort statewide data, middle school performance is stratified by gifted status and income as expected with gifted, higher income students scoring the highest on math and reading tests; however, there is no differential increase over time by income. All students, regardless of gifted status or income, increase their test outcomes at about the same rate over middle school in the actual data.

Since the proportions of gifted students by race/ethnicity and income are not representative of the larger school population across North Carolina middle schools, propensity score matching is used to create equivalent gifted and not-gifted groups for analysis. Again, the data presented is based on the matched middle school cohorts and therefore includes students with similar propensities to be in gifted programming. Hierarchical linear growth models are run to enable analysis of the impact of gifted programming, race/ethnicity and income over time on student outcomes.

4.2 Hierarchical Linear Model Results

The Full 2 model is selected as the hierarchical linear model of analysis for math and reading growth models. As previously explained, this is due to the increased explained variance and significance of additional covariates. Table 24 below is pulled from the table of both full hierarchical linear growth models.

Table 24: Full 2 Model – Fixed Effect Math and Reading

		Math	**Reading**
		(Full 2 Model)	
Intercept	Intercept *B00*	0.135*** (0.019)	0.253*** (0.016)
	Gifted MS *B017*	0.137*** (0.017)	0.197*** (0.018)
	Gifted G5 *B016*	0.100*** (0.015)	0.026 (0.018)
	Grade 5 Score *B018*	0.537*** (0.010)	0.464*** (0.009)
Slope	Time Slope Intercept *B10*	-0.062*** (0.008)	0.105*** (0.008)
	Gifted MS *B117*	0.012 (0.009)	0.081*** (0.009)
	Gifted G5 *B116*	-0.010 (0.009)	0.015 (0.009)
	Grade 5 Score *B118*	-0.014** (0.005)	-0.227*** (0.004)

Robust standard error in parenthesis, *p<.05, **p<.01, ***p<.001

Gifted middle school students have higher initial test scores than not-gifted students in both subjects. While previous performance appears to account for some of this difference, as shown by the decrease in the intercepts when the grade 5 score is added, there is still an unexplained gap between the test scores of not-gifted middle school students, gifted middle school students, and previously gifted students at the end of eighth grade. Students participating in gifted programming perform better in math and reading than students not in gifted programming; yet gifted students have significant gains only in reading performance over time when Grade 5 Score is included in the model. The coefficients on Gifted MS and GiftedG5 in the equation for math imply that there is no increase in test score over time, with or without the fifth grade test score in the model.

The gifted middle school dummy (Gifted MS) is the central variable of interest in the models as the differential impact of gifted programming is the question of interest. Students who are gifted in middle school perform about one tenth of a standard deviation higher in math and about one fifth of a standard deviation higher in reading than those not-gifted in middle school in the eighth grade, when comparing students of the same race/ethnicity, poverty status, parent education level, gender, English proficiency, disability status, age, previously gifted status, and fifth grade test scores. Both of these math and reading estimates are highly significant. Math performance is not significantly impacted over time for gifted middle school students, yet being gifted in middle school increases the average rate of growth in reading by 0.08 standard deviations compared to students who were not-gifted in middle school. Math students who were gifted in the fifth grade score one tenth of a standard deviation higher than students who were not previously gifted, holding all other variables constant. However gifted status in the fifth grade only significantly increase final math outcomes, reading performance at the end of middle school nor math or reading growth over time is significantly impacted by being gifted in elementary school.

The standardized test score from fifth grade has a large impact on both models. In reading, the test score is about one half of a standard deviation higher (0.46) for a one standard deviation increase in grade five score holding all other variables constant. This effect is just over half of a standard deviation for math (0.54). When examining growth, the previous test score has a negative effect in both subjects. In other words the higher the fifth grade test scores, the smaller the increase in middle school test scores. Although the coefficients are much smaller (0.01 and 0.23 standard deviations, for math and reading respectively), the performance of students decreases as the fifth grade score increases by one standard deviation.

In Figures 15 and 16, the predicted outcomes of gifted programming on student math and reading achievement comparing gifted students to not-gifted students are displayed graphically. The three years of middle school, 2004-2005 to 2006-2007, are shown below with separate lines for gifted, not-gifted, and all students.

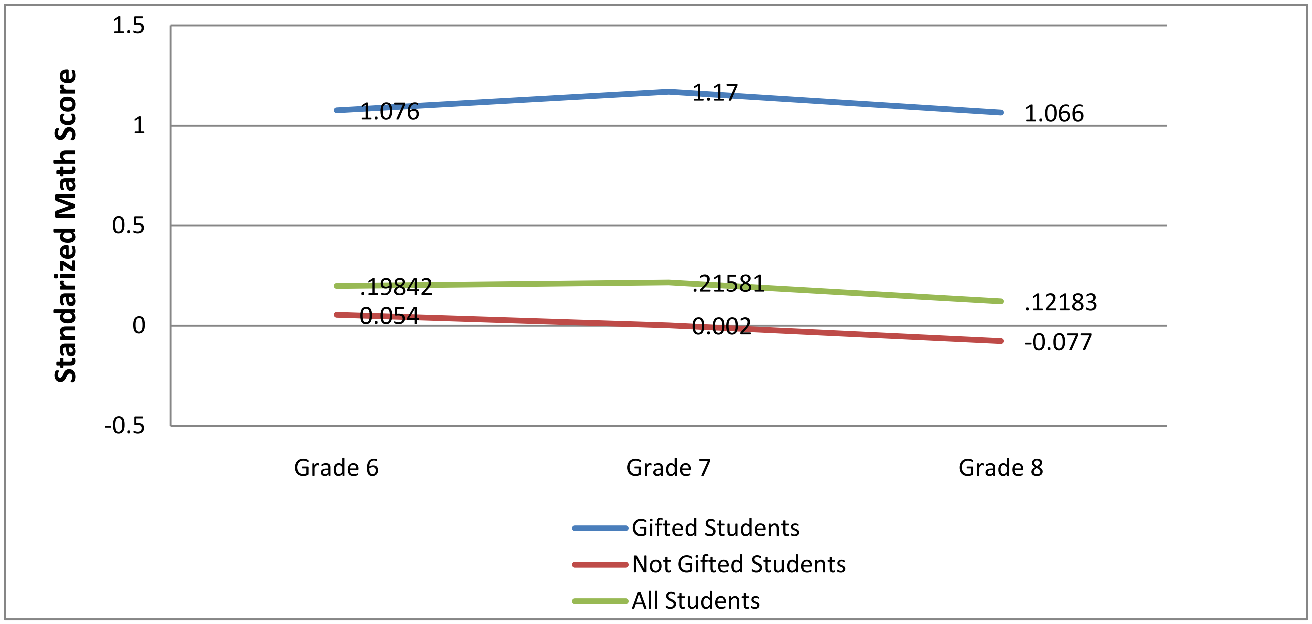


Figure 15: Predicted Outcomes of Gifted and Not-Gifted Students – Math

There is not a large change in the predicted test scores of math students over time. There is an initial gap between gifted and not-gifted students, however the change in test scores over time are largely the same. The slight hike in the seventh grade score is due to the math test changing in the 2005-2006 school year; test scores were re-scaled for the sixth grade to adjust for this in the analysis, however the test was different in grades seven and eight. The slopes are comparable for the three groups of students.

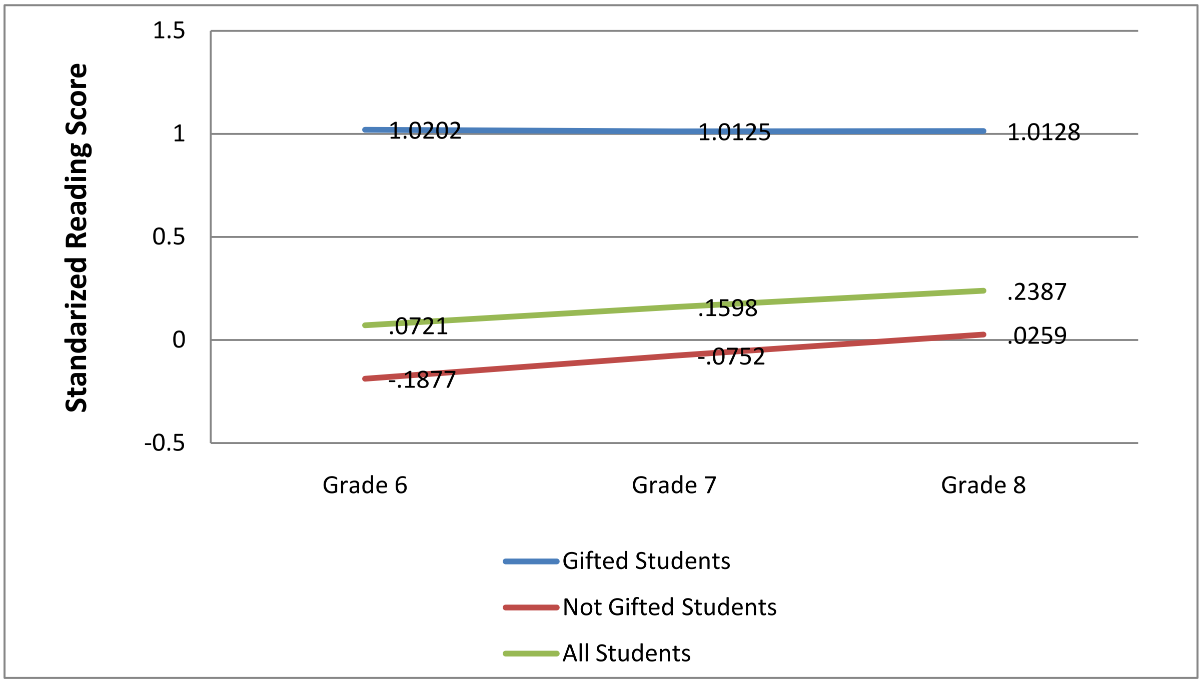


Figure 16: Predicted Outcomes of Gifted and Not-Gifted Students – Reading

For the reading test, there is again an initial gap between gifted and not-gifted reading students, yet over time gifted student reading performance is predicted to be quite flat. Gifted reading students do not increase in performance, however there is a positive, much steeper slope for both not-gifted and all students. It is clear that gifted students are not gaining at a higher rate than not-gifted students

Gifted students are expected to perform better than not-gifted students, which they do. However in the HLM predicted outcomes, gifted student performance is not predicted to increase greatly over time at a higher rate than not-gifted students as hypothesized. The slopes of test scores of math and reading students comparing not-gifted and all students are actually similar; yet, not all predicted rates of change are significant. Of the key variables listed in Table 24, only gifted middle school status for reading and black and free lunch status coefficients for both reading and math are significant at a .01 level or higher in the HLM analyses. Being previously labeled gifted does not significantly impact predicted growth of students in math or reading. Although consistent, as evident in the sub-groups below, these rates of change over time are not all significant.

Growth rates of gifted and not-gifted student outcomes are also compared graphically by race/ethnicity and income levels in Figures 17 to 20 below; the predicted outcomes align with the overall performance of gifted and not-gifted populations.

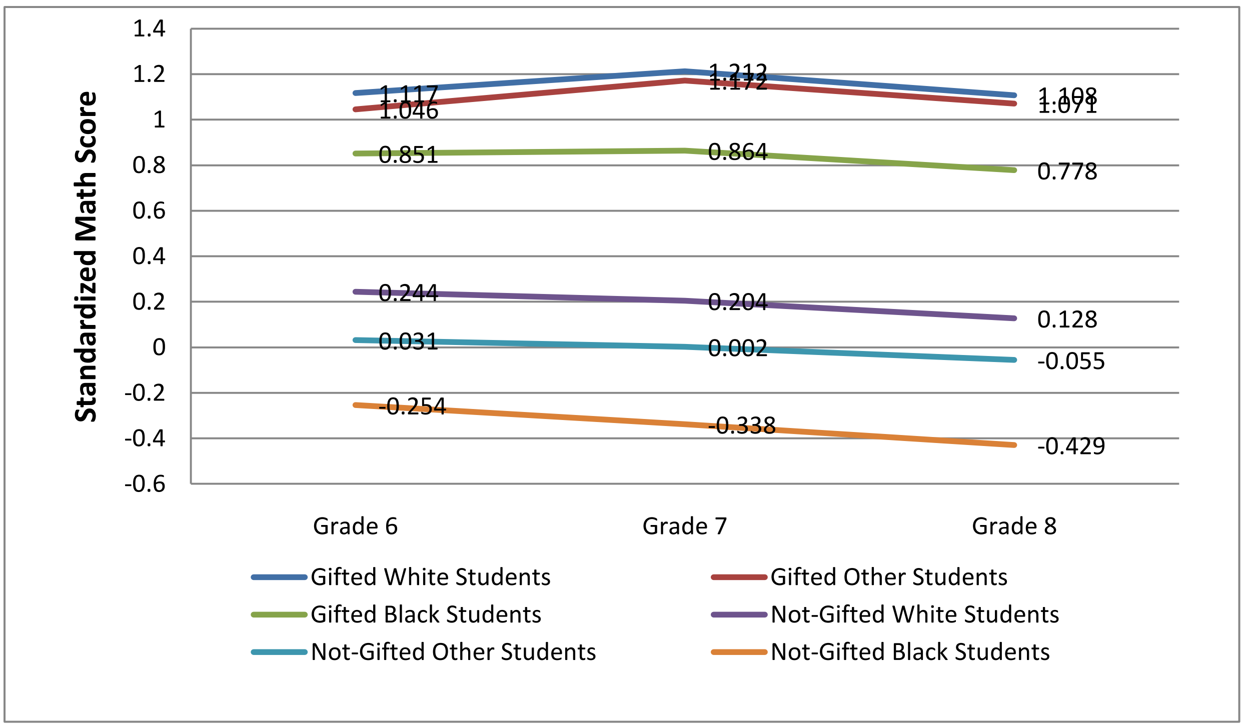


Figure 17: Predicted Outcomes of Students by Race/Ethnicity – Math

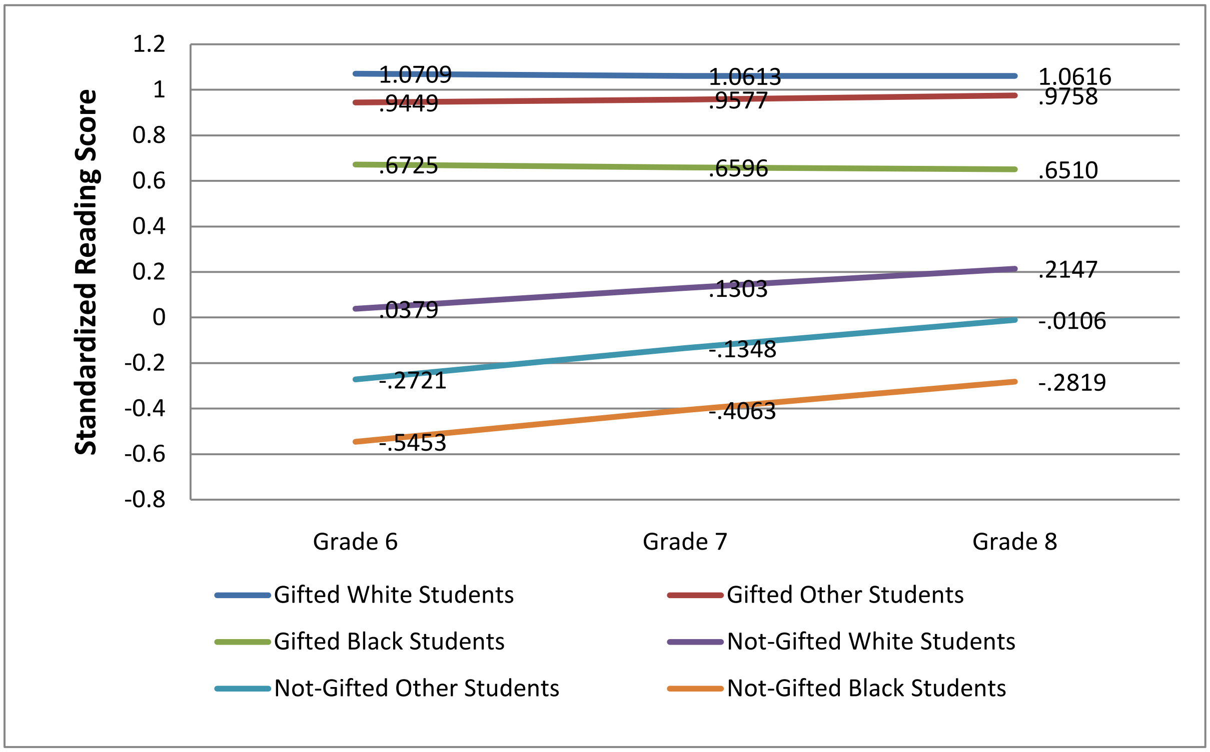


Figure 18: Predicted Outcomes of Students by Race/Ethnicity - Reading

When examining performance by race/ethnicity, black gifted students initial predicted performance is lower than both white and other gifted students. These findings are consistent with those of Cornell et al. (1995) that minority gifted students perform higher than not-gifted minority students, yet their scores are below those of their white gifted peers. The same gaps are evident for not-gifted students; black not-gifted students have lower scores than other students and white not-gifted students. In both reading and math, the performance of white and other gifted students is quite close, with a larger gap between black gifted students. This is not true with not-gifted students; the gaps are evenly spread across the groups, and remain so over time. While gifted student reading performance is flat and math performance flattens out over time, the predicted slope of math scores for not-gifted students is negative. These students are performing worse over time, and black students fare the worst. The reading performance of not-gifted students increases over time with even gaps across racial/ethnic groups.

Similar gaps and patterns are found in predicted outcomes by poverty status. Higher income students perform the highest initially followed by reduced lunch and free lunch students.

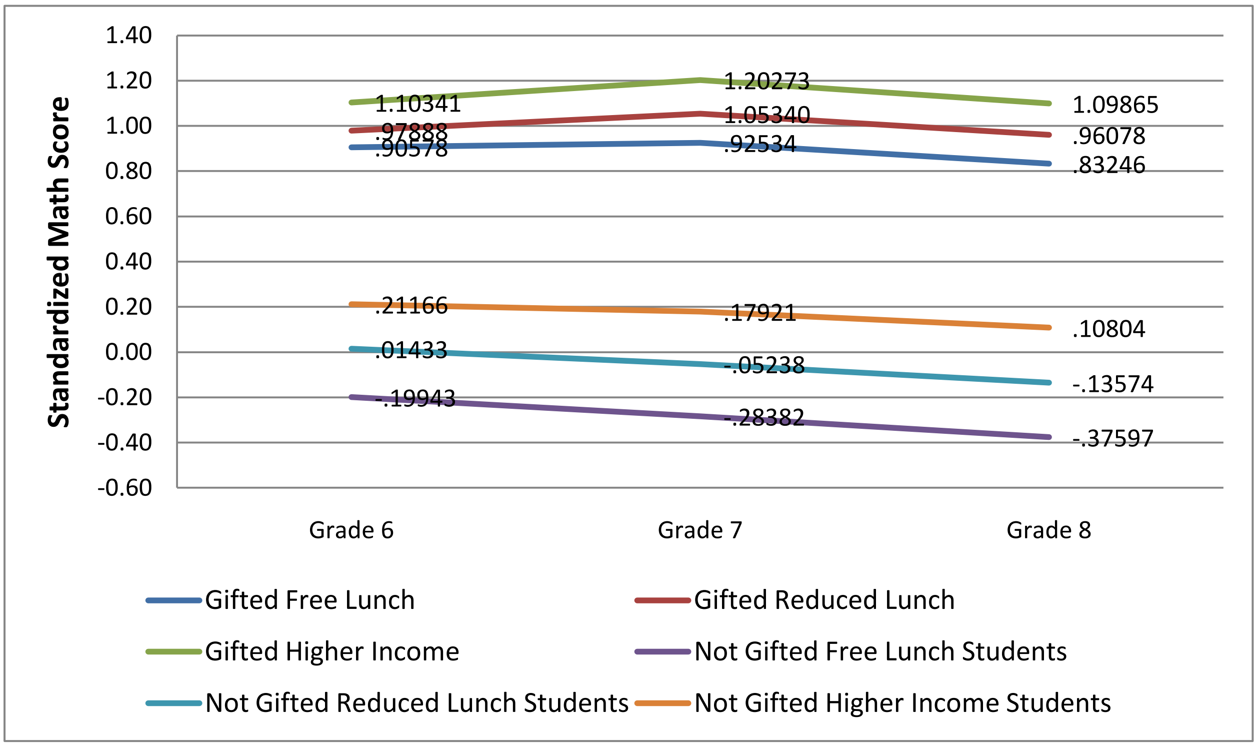


Figure 19: Predicted Outcomes of Students by Income – Math

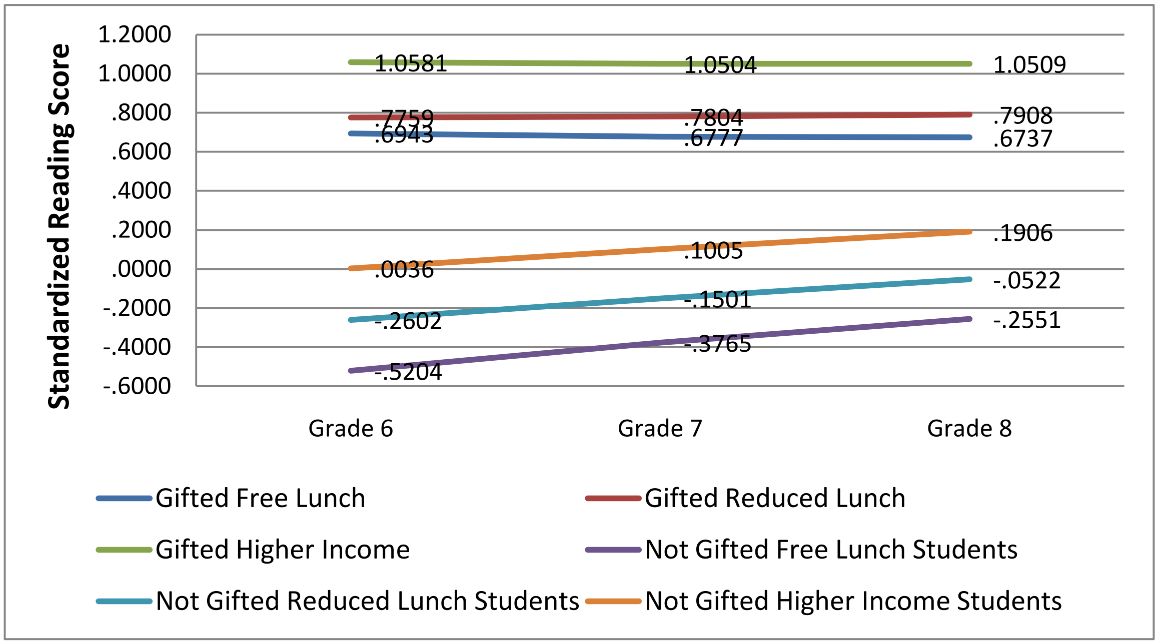


Figure 20: Predicted Outcomes of Students by Income - Reading

The slopes within these gifted and not-gifted groups are again very similar. Math not-gifted students have negative slopes signifying lower test scores over time, while performance of not-gifted reading students is again predicted to increase over time.

It is interesting to note the gains in predicted reading scores for not-gifted students throughout middle school compared to all other student populations. Not-gifted reading students make academic gains when examining all students and also when race/ethnicity groups and income levels are broken out. The results of not-gifted black and lower income reading students have greater gains than their gifted counterparts. However, the subgroups are still stratified. Black and lower income not-gifted students do not have larger gains than their white or higher income not-gifted peers. For instance, the trajectory of white not-gifted students remains higher than that of black not-gifted reading students. However, the slope for black and free lunch not-gifted students are increasing at a faster rate than for black and free and reduced lunch gifted students. Change in performance over time remains stratified by gifted status, race/ethnicity, and poverty status.   Math not-gifted performance decreases over time, however test score losses remain stratified. . It is important to note that standardized test scores are used, so although gifted performance appears quite flat these students are actually maintaining their average spot on the distribution of test scores. Not gifted students are predicted to increase in performance over time but that result over middle school equates to about where gifted students performed in fifth grade.

The reading findings are consistent with a recent study on the impact of NCLB on high-achieving student performance. In this two part study, Loveless (2008) examines gains of high and low achievers using National Assessment of Educational Progress (NAEP) data. A National Teacher Survey is also analyzed by Farkas and Duffett (2008). This time period analyzed is consistent with the years used in this study. Both parts of the recent NCLB study conclude that lower achieving students make stronger progress than high achievers. In the survey, teachers admit to focusing more attention on struggling learners. The top ten percent of fourth and eighth grade students make minimal gains in math and reading (five and three points, respectively) on NAEP since 2000. However students in the bottom ten percent show solid progress (13 and 16 points, respectively). Only 23 percent of teachers surveyed stated that "academically advanced" students were a top priority; 60 percent of teachers felt struggling students were a top priority.

Wherever students enter North Carolina middle schools, their academic growth is stratified by race/ethnicity and income raising significant questions regarding persistent inequities; performance is higher for those participating in gifted programming, yet the gaps remain. Although large improvements over time are not found, as expected, the differences in student groups remain constant through middle school, placing students on different pathways to high school. Ramifications arguably extend through high school and strongly influence graduation, post-secondary education, and workforce readiness.

Even though the gifted middle school coefficient does not have a large impact on student performance over time, other variables in the model impact outcomes in expected ways. Table 25 details the findings of all variables used in the analysis.

Table 25: HLM Math and Reading Results (Full 2)

		Math Coefficients		**Reading Coefficients**	
		Intercept	Slope	Intercept	Slope
School Level	Black	-0.001* (0.001)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	Other	0.001 (0.001)	0.001 (0.000)	0.003** (0.001)	0.001** (0.000)
	Free Lunch	-0.003** (0.001)	-0.001* (0.000)	-0.005*** (0.001)	-0.003*** (0.000)
	Reduced Lunch	-0.003 (0.002)	-0.002* (0.001)	-0.001 (0.002)	-0.000 (0.001)
	Teacher Turnover	0.003* (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
	School Size	0.000 (0.152)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000
Student Level	Black	-0.102*** (0.018)	-0.018* (0.009)	-0.139*** (0.017)	-0.069*** (0.009)
	Other	0.062** (0.018)	0.012 (0.009)	0.032 (0.019)	0.010 (0.009)
	Free Lunch	-0.086*** (0.017)	-0.020* (0.008)	-0.069*** (0.016)	-0.031*** (0.008)

Table 25 Continued

	Reduced Lunch	-0.028 (0.025)	-0.005 (0.011)	-0.043* (0.020)	-0.021* (0.010)
	Free Reduced Lunch Missing	-0.130*** (0.022)	-0.026* (0.011)	-0.039 (0.025)	-0.019 (0.012)
	Parent Less than High School	-0.010** (0.033)	-0.030* (0.015)	-0.080* (0.034)	-0.034* (0.016)
	Parent Some College	0.034* (0.015)	0.003 (0.008)	0.051** (0.014)	0.020** (0.007)
	Parent College Grad	0.119*** (0.014)	0.017* (0.007)	0.137*** (0.014)	0.059*** (0.007)
	Parent Education Missing	0.114** (0.038)	0.005 (0.024)	0.099 (0.053)	0.035 (0.026)
	Male	-0.020 (0.010)	-0.003 (0.005)	-0.022* (0.010)	-0.003 (0.005)
	Was LEP	0.094 (0.053)	0.018 (0.022)	0.030 (0.032)	0.014 (0.016)
	Is LEP	-0.033 (0.102)	0.035 (0.040)	-0.203** (0.057)	-0.093*** (0.025)
	Disability	-0.050 (0.042)	0.011 (0.020)	-0.052 (0.038)	-0.017 (0.019)
	Underage	0.055 (0.042)	0.037 (0.023)	0.009 (0.041)	0.007 (0.020)
	Overage	-0.089*** (0.018)	0.002 (0.009)	-0.048** (0.015)	-0.024** (0.007)
	Gifted Grade 5	0.100*** (0.015)	-0.010 (0.012)	0.026 (0.018)	0.015 (0.009)
	Gifted MS	0.137*** (0.017)	0.012 (0.009)	0.197*** (0.009)	0.081*** (0.009)
	Grade 5 Test Score	0.537*** (0.010)	-0.014** (0.005)	0.464*** (0.009)	-0.227*** (0.004)
Time	Intercept/Time Slope	0.135*** (0.019)	-0.062*** (0.008)	0.253*** (0.016)	0.105*** (0.008)

Robust standard error in parenthesis, *p<.05, **p<.01, ***p<.001

When looking at the final outcomes of eighth grade students, race/ethnicity has differing impacts on math and reading performance. Black students perform significantly lower than their white peers in both math and reading. However, for math only, other

(Hispanic, American Indian, Asian, or multiracial) students perform significantly higher than white students. The other race/ethnicity category does not significantly impact reading performance when compared to white students. Free lunch status significantly and negatively impacts outcomes for both subjects as compared to those who pay full price lunch. The impact of reduced lunch students is significant for reading scores only, suggesting there may be a line of demarcation of income level on student test performance between free and reduced lunch student designations. While the coefficient on “free lunch missing” is significant for math performance, it does not significantly impact reading outcomes.

Regarding parental education levels, compared to parents with a high school diploma (which is the reference group), parents with less than a high school education significantly decrease student outcomes in reading and math and the children of parents with some college education perform higher than those with no college experience. Having a parent who is a college graduate also positively impacts performance in both subjects. Finally, the coefficient for parent education data missing is significantly positive for math but not for reading.

Gender and limited English proficiency (LEP) both significantly impact reading outcomes, controlling for other characteristics. Male students perform worse than females and current LEP students perform worse than students proficient in English. Not surprisingly, students who are overage have lower math and reading test scores than students whose birthdays fall within the state cut-off to enter Kindergarten when this cohort first began school; this proxy for retention negatively impacts outcomes as expected.

Differences between school level characteristics are also examined in these HLM models. Having a larger percentage of free lunch students decreases average student performance in both reading and math. As the student population of free lunch students increases within a school, test scores in math and reading decrease. The coefficient on the percentage of students in the other racial/ethnicity category is positive and significant only in the reading model, while black school percentages only significantly negatively impact math performance.

Variables that significantly impact final reading test scores similarly affect estimates of growth in reading test scores throughout middle school. The only difference is that test scores for males do not increase at a faster rate than females when comparing similar students.

There are more differences in the gains of math outcomes over time when compared to their final middle school outcomes. Although eighth grade scores are significantly different, the growth rates of other students, those with missing parent education data, overage, gifted in the fifth grade, and gifted in middle school are no longer significant when compared to the growth rates of their respective peers. These characteristics significantly impact final differences in student math performance, but do not significantly impact student gains over time. The standardized test score from fifth grade remains highly significant; an increase in this score by one standard deviation results in a decrease in the average growth in math test scores by 0.01 each year.

The overall change in performance for students over the three years is -0.12 standard deviations in math and 0.21 standard deviations in reading; math performance decreases over time at a highly significant level (99%). The gain in reading for gifted middle school students is an 0.16 standard deviations; however, the gifted middle school slope is not significant for math. Therefore the gain for gifted reading students each year is less than one tenth of a standard deviation larger than not-gifted students while gifted math students do not experience any additional gain in outcomes. There are statistically significant differences for final test scores and changes across time for all students in reading and math.

To better understand the small effect of gifted programming, the HLM models are also run considering potential differences in the distribution of "ability" among gifted students. Instead of one gifted dummy variable from fifth grade (Gifted G5 = 1 if Gifted in fifth grade), students are divided into the highest and lowest gifted performers on Gifted G5. Fifth grade math and reading test scores are used as a proxy for ability and two sets of variables are created (Gifted G5 Top and Gifted G5 Bottom); these new variables are based on approximates of the top and bottom ten percent of gifted students and the top and bottom top twenty percent of gifted students. Table 26 below presents the HLM results for both math and reading outcomes when the ability variables are included.

Table 26: Model Comparisons – Fixed Effect Math and Reading by "Ability"

		Math		Reading	
		10 Percent	20 Percent	10 Percent	20 Percent
Intercept	Intercept	0.158*** (0.019)	0.143*** (0.018)	0.039*** (0.016)	0.059*** (0.015)
	Gifted MS	0.208*** (0.012)	0.205*** (0.012)	0.287*** (0.012)	0.286*** (0.012)
	Gifted G5 Top	-0.004 (0.015)	0.144*** (0.015)	-0.062** (0.015)	0.052 (0.016)
	Gifted G5 Bottom	-0.182* (0.078)	-0.369*** (0.049)	-0.262*** (0.049)	-0.320*** (0.050)
	Grade 5 Score	0.531*** (0.014)	0.458*** (0.014)	0.539*** (0.016)	0.465*** (0.016)
Slope	Time Slope Intercept	-0.055*** (0.008)	-0.073*** (0.008)	-0.030** (0.009)	-0.033*** (0.008)
	Gifted MS	0.004 (0.006)	0.003 (0.006)	0.008 (0.007)	0.008 (0.007)
	Gifted G5 Top	-0.044*** (0.009)	0.006 (0.008)	0.008 (0.010)	0.001 (0.010)
	Gifted G5 Bottom	0.113*** (0.028)	0.148*** (0.023)	0.025 (0.028)	0.051* (0.023)
	Grade 5 Score	0.033*** (0.007)	0.007 (0.008)	-0.010 (0.009)	-0.004 (0.010)

The results suggest that students at the bottom end of the ability distribution have lower outcome measures. Gifted G5 Bottom, for both the lowest 10 and 20 percent of students, negatively impact the final difference in student test scores. Gifted students with the lowest previous ability are predicted to have significantly lower test scores than the middle distribution of gifted students. High ability gifted students (Gifted G5 Top) are predicted to have significantly higher test scores than the middle distribution only for the highest twenty percent in math. The top 10 percent of reading gifted fifth graders (Gifted G5 Top) are actually predicted to have significantly lower test scores than the middle distribution. Being a top reading gifted student results in a lower increase in performance over time than a similar student in the middle ability range. This suggests that there may be variation in test scores and rates of change of gifted students based on their previous ability.

When controlling for the highest and lowest ability gifted elementary school students, the predicted impact of gifted programming (Gifted MS) on their final status is still significantly positive across all groups. Gifted programming increases outcomes for all top and bottom gifted ability groups when compared to similar students in the middle ability gifted range. Yet as in previous models, this impact does not hold across time; the gifted middle school slope is not significant in any of the ability models.

Based on the time slope intercept, student performance is predicted to decrease over time in each model. In the math model, the test scores for the highest ability gifted students (10 %) decrease significantly over time while scores for the bottom, both 10 and 20 percent, significantly increase over time. There is also a significant increase for the bottom 20 percent of gifted students in reading, but not the bottom ten or either of the highest ability reading groups. This suggests that for most lower ability gifted students, gifted programming positively impacts performance over time. There is little to no difference in growth of the highest ability gifted students. There are variations in the expected outcomes of gifted students according to previous ability, however the gifted programming (Gifted MS) is only significant for the final performance of students; there is no significant effect over time. There may be differential impacts within gifted populations based on previous ability, however there is still no impact of middle school gifted programming in the gain scores over time. Enrollment in gifted program does not appear to increase student growth in reading or math outcomes.

It is clear there is a test score gap between gifted and not-gifted students. Gifted students have higher test scores than not-gifted students in both the descriptive statewide data and when controlling for student and school characteristics in the HLM cohort models. These outcomes are stratified by race/ethnicity and income, creating a question of equity. Since students in gifted programming have higher test scores and the gaps remain over time, gifted programming may lead to persistent gaps in attainment. North Carolina middle school students are matched in order to compare students with similar propensities to be gifted; therefore, the predicted outcomes are not simply a result of gifted students having different characteristics than not-gifted students. There may be systemic results from being labeled gifted and placed into a gifted program.

It is hypothesized that gifted students perform higher than not-gifted students which occurs in the test scores at each point in time. This holds true across race/ethnicity and income with black and lower income gifted students having lower test scores than white and higher income gifted students. However, the growth rates of gifted students do not increase at a higher rate than not-gifted students. There is actually a small effect found from being gifted. This unexpected finding could be a result of a mismatch with the gifted curriculum. It is argued that no single assessment can accurately measure both accountability and classroom instruction goals (Stiggens, 2003). The End of Grade test, the outcome measure, is based on the state's standard course of study. The curriculum of gifted students covers more as students are expected to utilize higher order thinking skills and are taught more challenging material.

Watanabe (2008) examined the impact of high stakes statewide accountability systems on equitable opportunities for children. After observing North Carolina middle school classrooms and interviewing teachers, he found that “students in the ‘academically gifted’ classes additionally received more opportunities to practice a wider range of reading and writing skills, engaged in more challenging instruction and assignments and received more written and immediate feedback on essay assignments than their peers in the ‘regular’ classes” (500). He also found that direct test preparation aligned with the End of Grade test was much more common in regular classes than gifted classes; approximately 45% of time in regular classes compared to just 15% in gifted classes. Table 26 below is taken from the Watanabe study.

Table 26: Percent of Observed Classroom Instruction by Level of Correspondence with Testing Demands and Format

	Level of Correspondence			
Class	High	Medium	Low	None
Regular	46.15%	23.08%	30.77%	0%
Academically Gifted	15.38%	61.54%	23.08%	0%

A low level of correspondence indicated that classroom instruction was tangentially related to the demands of the test. For example, learning to diagram sentences, a demand that neither the writing or multiple-choice tests required, might have helped students with their writing as they learned more about sentence formation.

Since much less time was spent on test preparation in gifted classes, these students were able to complete more advanced work. These findings are similar to previous studies (Oakes, 1985:76; Smith-Maddox & Wheelock, 1995) that examine the differences in learning environments across tracks. High stakes testing does not appear to increase opportunities for children, but encourage explicit test preparation, especially for students in the regular track (Watanabe, 2008). "Students of color and students from low socioeconomic backgrounds in the 'regular' track are thus shortchanged in opportunities to learn even in the new context of statewide accountability policies that policymakers purport to be part of the state's response to close the achievement gap" (Watanabe, 2008: 522).

CHAPTER FIVE: CONCLUSIONS AND IMPLICATIONS

5.1 Conclusions

This analysis quantifies gifted programming opportunities across North Carolina middle schools and how this differs by race/ethnicity and income. It is clear that black and lower income students are less likely to participate in gifted programming. Although black students are 31% of the middle school math student population from 2004-2005 to 2006-2007, they comprise just 12% of the gifted population. White students are 58% of the math population, however 79% of gifted students. The inverse is true for lower and higher income students; about 42% of students are lower income, yet just 16% of the gifted population compared to higher income students who are 55% of the total population and almost 80% of gifted students. Similar proportions are evident for reading middle school students as well. Black students are almost 30% of the reading population, yet only 10% of gifted students in contrast to over 80% of white students gifted in reading although white students comprise just 60% of the population. Finally, according to income, lower income students are 39% of the population versus just 14% of the gifted population and higher income students are 80% of the gifted population and 55% of the total population.

The results of the hierarchical linear growth models are not as expected. Although students in gifted programming have significantly higher test scores than their not-gifted peers, these test scores do not increase at a significantly faster rate over time for gifted math students and gains are smaller than expected for reading students. Thus while gaps in learning exist, I do not find that gifted learners are experiencing much larger increases in test scores throughout middle school than not-gifted learners as hypothesized. The gaps remain across both race/ethnicity groups and income. For instance black gifted students have higher test scores than black and white not-gifted students; however, their scores are not as high as white gifted students. The same is true for income levels; both gifted and not-gifted student outcomes are stratified by higher income, reduced lunch and free lunch designations. This suggests that these initial differences resulting from gifted placement in elementary school persist throughout middle school, ultimately impacting the final scores of students as they enter high school. Although gifted programming does not greatly increase the growth of student performance, it does appear to maintain gaps in test scores over time. It is important to note that although gifted middle school students are not making large gains, there performance is maintained over time and is higher than not-gifted students throughout middle school. Outcomes are stratified by gifted programming and the race/ethnicity and income of middle students.

5.2 Policy Implications

There are several potential reasons that students in gifted programming are not increasing test score performance over time. The test could simply be a mismatch to what gifted students are expected to do in the classroom. It is argued that no single assessment can accurately measure both accountability and classroom instruction goals (Stiggens, 2003). The End of Grade tests used in this analysis are based on the standard course of study of the state; they are not designed to be a complete measure of gifted student curriculum and learning. Additionally, Watanabe (2008) observes North Carolina middle school classrooms and notes the wider range of skills captured in gifted than regular classes. He also finds regular classes utilize more direct test preparation aligned with the End of Grade test than did gifted classes. The increasing federal accountability demands have created an environment of high stakes testing that appears to encourage explicit test preparation for regular track classes in particular (Watanabe 2007). Additional measures of learning and student performance may be necessary to further estimate changes in growth over time resulting from gifted programming.

The learning environment of “regular” classes, arguably created as a result of high-stakes accountability tests, is a significant concern for student achievement. The extent to which high stakes testing negatively impacts or shortchanges learning outcomes for students is an area of research that merits further examination. There are findings of test-preparation at the expense of learning and rigorous curriculum in urban and lower income areas in particular (Delpit, 2003; McNeil, 2000; McNeil & Valenzuela, 2000; Scheurich, Skrla, & Johnson, 2000). Yet, the reach may be further; perhaps there is no significant growth in gifted learners because the majority of teachers are in fact more concerned with ensuring lower performers reach proficiency as suggested in the National Teacher Survey (Farkas & Duffett, 2008). They find that lower achieving students actually make stronger progress than higher achievers on the National Assessment of Educational Progress (NAEP). Teachers admit to focusing more attention on struggling learners than on those who are “academically advanced.”

Regardless of the merits of standardized tests, the central question of equity remains. Students finish middle school at different performance levels according to whether or not they participated in gifted programming, their race/ethnicity, and their poverty status. These factors cannot be determinants for student outcomes. Although

End of Grade tests are not a complete measure of student learning, they do quantify what information a child knows based on a standardized measure. Since a cohort of students are matched in this analysis through propensity score matching, differences in test score outcomes are not simply attributable to inherent differences in students. If regular classes spend more time on test preparation and still fare worse than gifted students, there is a structural problem that requires further attention. A more comprehensive examination of gifted compared to not-gifted classroom learning environments is needed. As concluded in existing gifted literature, the impact of gifted programming or ability grouping may be due to the degree of curriculum changes that occur within learning environments and not the grouping mechanism itself.

Any program that results in stratified outcomes for children over time needs careful examination to ensure all children receive an equitable learning experience and opportunity to reach their full potential.

APPENDIX A: LOGISTIC REGRESSIONS

Math

Logistic regression	Number of obs	=	50988
	LR chi2(15)	=	20674.13
	Prob > chi2	=	0.0000
Log likelihood = -13834.519	Pseudo R2	=	0.4277

ex aig05	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ma scoreg5	1.318197	.0040921	88.99	0.000	1.310201	1.326242
black	.7114459	.0375389	-6.45	0.000	.6415476	.7889599
other	.7623498	.043355	-4.77	0.000	.6819402	.8522407
frlnch	.4582669	.0214678	-16.66	0.000	.4180647	.5023351
redlnch	.542733	.0358129	-9.26	0.000	.4768905	.6176662
male	.7730487	.0238024	-8.36	0.000	.7277766	.821137
per below ~s	.9035396	.1355772	-0.68	0.499	.6733229	1.21247
per adv de~s	.9926486	.0017794	-4.12	0.000	.9891673	.9961423
per 0 3 yr~p	.9965595	.0029945	-1.15	0.251	.9907075	1.002446
per 4 10 y~p	1.000236	.0020719	0.11	0.909	.9961831	1.004305
total_tch~04	1.008982	.0031923	2.83	0.005	1.002744	1.015258
adm04	.99906	.000168	-5.59	0.000	.9987307	.9993894
per freere~h	.9980884	.0006895	-2.77	0.006	.9967379	.9994408
per black~04	1.012254	.0009769	12.62	0.000	1.010341	1.01417
per other~04	1.009109	.0018308	5.00	0.000	1.005527	1.012704

Reading

Logistic regression	Number of obs	=	41917
	LR chi2(15)	=	14467.95
	Prob > chi2	=	0.0000
Log likelihood = -13259.333	Pseudo R2	=	0.3530

ex_aig05	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
rd_scoreg5	1.282988	.0042165	75.82	0.000	1.274751	1.291279
black	.5886693	.0311872	-10.00	0.000	.5306099	.6530815
other	.9736308	.0549971	-0.47	0.636	.8715912	1.087617
frlnch	.4034186	.0195474	-18.73	0.000	.3668694	.443609
redlnch	.5119435	.0355798	-9.63	0.000	.4467494	.5866514
male	1.070415	.0329706	2.21	0.027	1.007706	1.137027
per below ~s	.9310144	.1219023	-0.55	0.585	.720285	1.203396
per adv de~s	.9968188	.0017874	-1.78	0.076	.9933217	1.000328
per 0 3 yr~p	.9967474	.0033539	-0.97	0.333	.9901954	1.003343
per 4 10 y~p	.9982399	.0021351	-0.82	0.410	.9940639	1.002433
total_tch~04	1.011124	.0035839	3.12	0.002	1.004124	1.018173
adm04	.9992847	.0001759	-4.06	0.000	.99894	.9996296
per freere~h	.9987534	.0004225	-2.95	0.003	.9979257	.9995817
per black~04	1.012045	.000941	12.88	0.000	1.010202	1.013891
per_other~04	1.00398	.0017159	2.32	0.020	1.000623	1.007349

APPENDIX B: Reading Epanechikov Matched Sample

Variables	**Absolute Value of Differences in Means: All Not-Gifted Students and Gifted Students**	**Absolute Value of Differences in Means: Matched Students and Gifted Students**	**Reduction in Bias**
Reading score	9.8868	0.1988	97.99%
Gifted (Treatment Variable)	1	1	0.00%
White	0.2614874	0.0017672	99.32%
Black	0.2252318	0.0005577	99.75%
Other	0.0362555	0.0023247	93.59%
Male	0.0232847	0.0023223	90.03%
free lunch	0.3025006	0.0027211	99.10%
reduced lunch	0.04997	0.0012404	97.52%
full pay	0.3524705	0.0014806	99.58%
parent less than high school	0.0922791	0.0131177	85.78%
parent high school only	0.2851235	0.1006256	64.71%
parent some college	0.0479505	0.0391735	18.30%
parent college grad	0.4277215	0.1556787	63.60%
parent edu missing	0.0023683	0.0027619	-16.62%
tchrsstayed04	1.92834	0.02174	98.87%
tchr retention 04	0.82919	0.23006	72.25%
per nbpts 04	1.422191	0.315126	77.84%
per below bachelors	0.0043369	0.0003229	92.55%
per advanced degrees	1.21045	0.09854	91.86%
percent 0-3 years experience	0.83054	0.0063	99.24%
percent 4-10 years experience	0.91574	0.10433	88.61%
percent 11+ years experience	2.03408	0.74597	63.33%
total teachers 04	1.94027	0.12155	93.74%
school size	38.3959	4.4122	88.51%
percent free/reduced lunch	12.34681	0.11304	99.08%
black mean	4.62829	0.23744	94.87%
white mean	5.4625	0.14304	97.38%
other mean	0.83422	0.09439	88.69%
Propensity score	0.3515738	0.0044966	98.72%

REFERENCES

Ainsworth-Darnell, J. W. and D. B. Downey (1998). Assessing the oppositional culture explanation for racial/ethnic differences in school performance. *American Sociological Review*, *63*(4), 536-553.

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*(4), 623-645.

Baker, B. D. (2001). Gifted children in the current policy and fiscal context of public education: A national snapshot and state-level equity analysis of Texas. *Educational Evaluation and Policy Analysis*, *23*(3), 229-250.

Baldwin, A. Y. (2005). Identification concerns and promises for gifted students of diverse populations. *Theory into Practice*, *44*(2), 105.

Barton, P.E. (2003). Parsing the achievement gap: Baselines for tracking progress. Princeton, NJ: Educational Testing Service.

Baser, O. (2006). Too Much Ado about Propensity Score Models? Comparing Methods of Propensity Score Matching. *Value in Health, 9*(6), 377-385.

Becker, S. & Ichino, A. (2002). Estimation of average treatment effects. *Stata Journal, 2*(4), 358-377.

Betts, J. R. & Shkolnik, J. L. (2000). Key difficulties in identifying the effects of ability grouping on student achievement." *Economics of Education Review, 19*(1), 21-26.

Bishop, K., Dudley, P. M., Mihaly, K., & Murphy, A. (2005). The lasting impact of teachers: An analysis of teacher race and student achievement. Working Paper: Duke University.

Blackburn, M. (1999). Studying privilege in a middle school gifted classroom. In J. Allen (Ed.), *Class Actions: teaching for social justice in elementary and middle school.* (pp. 72-83). New York: Teachers College Press.

Boaler, J. (2003). When learning no longer matters: standardized testing and the creation of inequality. *Phi Delta Kappan*, *84*(7), 502.

Bowles, S. & Gintis, H. (2002). Schooling in Capitalist America revisited. *Sociology of Education*, *75*(1), 1-18.

Brody, L. E. & Mills, C, J. (2005). Talent search research: what have we learned?. *High Ability Studies*, 16(1), 97.

Brown, E., Avery, L., VanTassel-Baska, J., Worley li, B. I, & Stambaugh, T. (2006). Legislation and policies: Effects on the gifted. *Roeper Review, 29(*1), 11.

Bryk, A. S., Nagaoka, J. K., & Newmann, F.M. (2000). Chicago classroom demands for authentic intellectual work: Trends from 1997-1999. Chicago: Chicago Consortium on Chicago School Research.

Carter, P. (2005). *Keeping It Real: School Success Beyond Black and White*. New York: Oxford University Press.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review*, *24(*4), 377-392.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources, 41*(4), 778-820.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review, 26*(6), 673-682.

Cook, T. D., Shadish, W. R., & Wong, V.C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management. 27*(4), 724-750.

Cornell, D. G., Delcourt, M. A., Goldberg, M. D., & Bland, L. C. (1995). Achievement and self-concept of minority-students in elementary-school gifted programs. *Journal for the Education of the Gifted, 18(*2), 189-209.

Darling-Hammond, L. (2007). A marshall plan for teaching. *Social Policy*, *37(*3/4), 41-44.

Darity, W, Castellino, D., Tyson, K., Cobb, C. & McMillen B. (2001). *Increasing opportunity to learn via access to rigorous courses and programs: One strategy for closing the achievement gap for at-risk and ethnic minority students*. Raleigh: North Carolina Department of Public Instruction.

Davidson Institute for Talent Development. (2010). Retrieved September 19, 2010. http://www.davidsongifted.org/db/

Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics*, *86*(1), 195-210.

Delcourt, M. A. B., Cornell, D. G., & Goldberg, M. D. (2007). Cognitive and affective learning outcomes of gifted elementary school students. *Gifted Child Quarterly, 51*(4), 359-381.

Delpit, L. (2003). Educators as "Seed People" growing a new future. *Educational Researcher, 32*(7), 14-21.

Downey, D. B. & Ainsworth-Darnell, J, W. (2002). The search for oppositional culture among black students. *American Sociological Review*, *67*(1), 156-164.

Downey, D. B. & Pribesh, S. (2004). When race matters: Teachers' evaluations of students' classroom behavior. *Sociology of Education*, *77*(4), 267-282.

Ehrenberg, R. G., Goldhaber, D., & Brewer, D. (1995). Do teachers' race, gender, and ethnicity matter? Evidence from the national educational longitudinal study of 1988. RAND Reprints. *Industrial and Labor Relations Review, 48*(3), 547-561.

Elhoweris, H., Mutua K., Alsheikh, N, & Holloway, P. (2005). Effect of children's ethnicity on teachers' referral and recommendation decisions in gifted and talented programs. *Remedial and Special Education, 26*(1), 25-31.

Farkas, S. & Duffett, A. (2008). High-achieving students in the era of NCLB: Part 2 results from a teacher survey. Washington, DC.: Thomas B. Fordham Institute.

Ferguson, R. (2002). Addressing racial disparities in high-achieving suburban schools. *NCREL Policy Issues 1*, 3–11.

Fiedler, E. D. & Lange, R. E. 1993). In search of reality: Unraveling myths about, tracking, ability grouping and the gifted. *Roeper Review*, *16*(1), 4.

Fiedler, E. D., Lange, R. E., & Winebrenner, S. (2002). In search of reality: Unraveling myths about, tracking, ability grouping and the gifted. *Roeper Review*, *24*(3), 108.

Ford, D. Y., Grantham, T. C., & Whiting, G. W. (2008). Another look at the achievement gap - Learning from the experiences of gifted black students. *Urban Education, 43*(2), 216-239.

Ford, D. Y., Harris III, J. J., Tyson, C. A., & Trotman, M. F. (2002). Beyond deficit thinking. *Roeper Review*, *2*(2), 52-58.

Fordham, S. & Ogbu, J, U. (1986). Black students', school success: Coping with the "burden of 'acting white.'" *Urban Review*, *18*(3), 176.

Fram, M. S., Miller-Cribbs, J.E., & VanHorn, L. (2007). Poverty, race, and the contexts of achievement: Examining educational experiences of children in the U.S. south. *Social Work, 52*(4), 309-319.

Gallagher, J. J. (2005). The role of race in gifted education. *Roeper Review*, *27*(3), 135.

Gardner, H., & Hatch, T. (1989). Multiple intelligences go to school: educational implications of the theory of multiple intelligences. *Educational Researcher*, *18*(8), 4-9.

Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* Cambridge: Cambridge University Press.

George, P., & Harrison, J. (2001). Representation of minority students in gifted and remedial programs and implications for closing the achievement gap in North Carolina middle schools. Durham: North Carolina Central University.

Gessner, S. L. (2008). The gifted express, now leaving on track 1. *Education Week*, pp. 28-28.

Gibson, M. A. & Ogbu, J. U. (Eds.). (1991). *Minority status and schooling: A comparative study of immigrant and involuntary minorities*. New York: Garland Publishing Inc.

Goetz, T., Preckel, F., Zeidner, M. & Schleyer, E. (2008). Big fish in big ponds: A multilevel analysis of test anxiety and achievement in special gifted classes. *Anxiety Stress and Coping*, *21*(2), 185-198.

Goldhaber, D. D., & Brewer, D. J. (1997). *Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity*. *Journal of Human Resources,* 32(3), 505-523.

Goldhaber, D. D. & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*. *22*(2), 129-145.

Hallinan, M. T. & Sorenson, A. B. (1977). The dynamics of learning: A conceptual model (Discussion Paper 444-77). Madison, WI: Institute for Research on Poverty.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). The market for teacher quality (NBER Working Paper No. 11154). Cambridge, MA: *National Bureau of Economic Research.*

Hanushek, E. A., Kain, J. F. & Rivkin, S. E. (2004). New evidence about "Brown v. Board of Education": The complex effects of school racial composition on achievement. (Discussion Paper no. 1284-04). Madison, WI: Institute for Research on Poverty.

Hawk, P. P., Coble, C. R., & Swanson, M. (1985). Certification: It does matter. *Journal of Teacher Education. 36*(3), 13-15.

Haycock, K. (2001). Closing the achievement gap. *Educational Leadership,* 58(6), 6-11.

Henry, G., Thompson C., Fortner F., Rickman, D. & Zulli-Lowe, R. (2008). The impact of disadvantaged student supplemental fund on high school student performance in pilot districts. Chapel Hill: Carolina Institute for Public Policy.

Henry, G. T., Fortner, C. K. & Thompson, C. L. (2010) Targeted funding for educationally disadvantaged students: a regression discontinuity estimate of the impact of high school achievement. *Educational Evaluation and Policy Analysis. 32*(2), 183-204.

Henry, G. & Yi, P. (2009, November). *Design matters: a within Study assessment of propensity score matching designs.* Paper presented at the Association for Public Policy Management, Washington, DC.

Hess, F. M. & Leal, D.L. (1997). Minority teachers, minority students, and college matriculation: A new look at the role-modeling hypothesis. *Policy Studies Journal,* 25(2), 235-248.

Johnsen, S. and VanTassel-Baska, J. (2006). National Standards for Teachers of Gifted and Talented Students. Swansboro, NC: North Carolina Association for the Gifted and Talented.

Klitgaard, R., & Hall, G. (1974). Are there unusually effective schools? *Journal of Human Resources, 74,* 90-106.

Kulik, J. A., & Kulik, C.-L. (1992). Meta-Analytic Findings on Grouping Programs. *Gifted Child Quarterly, 36*(2), 73-77.

LeTendre, G. Hofer, B, & Shimizu H. (2003). What is tracking? Cultural expectations in the U. S., Germany, and Japan. *American Educational Research Journal*, *40,* 43-89.

Lewis, A. (2003). *Race in the Schoolyard: Negotiating the Color Line in Classrooms and Communities*. New Brunswick,NJ: Rutgers University Press.

Loveless, T. (1999). Will tracking reform promote social equity? *Educational Leadership, 56*(7), 28-32.

Loveless, T. (2008). High-achieveing students in the era of NCLB: Part 1 An analysis of NAEP data. Washington, DC.: Thomas B. Fordham Institute.

Loveless, T., & Thomas, B. F. I. (2009). *Tracking and Detracking: High Achievers in Massachusetts Middle Schools.* Washington, DC: Thomas B. Fordham Institute.

Lucas, S. R. (1999). *Tracking inequality*. Stratification and mobility in American high schools. New York: Teachers College Press.

Ma, X. & Wilkins, J. L. M. (2002). The development of science achievement in middle and high school - Individual differences and school effects. *Evaluation Review*, 26(4), 395-417.

McBee, M. T. (2006). A descriptive analysis of referral sources for gifted identification screening by race and socioeconomic status. *Journal of Secondary Gifted Education, 17*(2), 103.

McClure, C. T. (2007). Ability grouping and acceleration in gifted education. *District Administration*, *43*(8), 24.

McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.

McNeil, L., & Valenzuela, A. (2000). *The Harmful Impact of the TAAS System of Testing in Texas: Beneath the Accountability Rhetoric*.

Meier, K.J. (1993). Latinos and representative bureaucracy: Testing the Thompson and Henderson hypotheses. *Journal of Public Administration Research and Theory, 3*(3), 393-414.

Meier, K. J., Doerfler, C., Hawes, D., Hicklin, A., & Rocha, R. (2006). The role of management and representation in improving performance of disadvantaged students: An application of Bum Phillips's Don Shula Rule. *Review of Policy Research*, *23,*5, 1095-1110.

Meier, K. J., Wrinkle, R. D., & Polinard, J. L. (1999). Representative bureaucracy and distributional equity: Addressing the hard question. *Journal of Politics, 61*(4), 1025-1039.

Morris, J. E. (2002). African American students and gifted education. *Roeper Review, 24*(2), 59-62.

Murnane, R.J. and Phillips, B.R. (1981). What do effective teachers of inner-city children have in common? *Social Science Research*, 10, 83-100.

Murnane, R. J., & Steele, J. L. (2007). What is the problem? The challenge of providing effective teachers for all children. *Future of Children, 17*(1), 15-43.

National School Lunch Program. (2007). Retrieved April 26, 2007, from http://www.fns.usda.gov/cnd/Lunch/AboutLunch/NSLPFactSheet.htm

North Carolina Department of Public Instruction. (2010). *Academically or intellectually* *gifted.* Retrieved November 2, 2010, from http://www.ncpublicschools.org/academicservices/gifted/

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven: Yale University Press.

Orfield, G., Frankenberg, E., & Garces, L. M. (2008). Statement of American social scientists of research on school desegregation to the U.S. Supreme Court in Parents v. Seattle School District and Meredith v. Jefferson County. *Urban Review, 40*, 96–136.

Peske, H.G., Haycock, K. & (2006). Teaching equality: How poor and minority students are shortchanged on teacher quality. Washington DC.: The Education Trust.

Pitts, D. W. (2005). Diversity, representation, and performance: Evidence about ethnicity in public organizations. *Journal of Public Administration Research and Theory, 15*(4), 323-339.

Pitts, D.W. (2007). Representative bureaucracy, ethnicity, and public schools: examining the link between representation and performance. *Administration and Society, 39*(4), 497-526.

Pfeiffer, S. I. and Jarosewich, T. (2007). An analysis of the standardization sample based on age, gender, race, and diagnostic efficiency. *Gifted Child Quarterly*, 51, 39-50.

Preckel, F., Zeidner, M., Goetz, T., & Schleyer, E. J. (2008). Female "big fish" swimming against the tide: the "big-fish-little-pond effect" and gender-ratio in special gifted classes. *Contemporary Educational Psychology. 33*(1), 78-96.

Ravallion, M. (2001). The mystery of the vanishing benefits: an introduction to impact evaluation. *The World Bank Economic Review, 15*(1), 115-140.

Robinson, A. (1990). Does that describe me? Adolescents' acceptance of the gifted label. *Journal for the Education of the Gifted, 13*(3), 245-255.

Robinson, A. & Clinkenbeard, P. R. (1998). Giftedness: An exceptionality examined. *Annual Review of Psychology, 49*, 117-139.

Roscigno, V. & Ainsworth-Darnell, J. (1998). Tracking matters. In J. P. McNamara (Ed.), *Perspectives in Social Problems* (pp. 158-165). Boulder: Coursewise Publishing.

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

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

Rubin, D. B. (1973). Matching to remove bias in observational studies. *Biometrics, 29*(1), 159-183.

Sadowski, M. (2006). The "Acting White" theory: What new teacher need to know. *Field Notes, 2*(1). Retrieved November 9, 2010 from http://www.bard.edu/mat/field-notes/archive.shtml?aid=349&pid=32

Scheurich, J. J., Skrla, L., & Johnson, J. F. (2000). Thinking carefully about equity and accountability. *Phi Delta Kappan, 82*(4), 293.

Schneider, A., & Ingram, H. (1993) Social construction of target populations: Implications for politics and policy. *American Political Science Review, 87*(3), 334-47.

Scientific Software International. (2010) Retrieved August 7, 2010 from http://www.ssicentral.com/hlm/example6-2.html

Selden, S. C. (1997). *The promise of representative bureaucracy: Diversity and responsiveness in a government agency*. Armonk, NY: ME Sharpe.

Seltzer, J. M. (1995). Creating healthy indoor environments - a road map for the future. *Occupational Medicine-State of the Art Reviews, 10*(1), 229-244.

Shore, B. M. & Delcourt, M. A. B. (1996). Effective curricular and program practices in gifted education and the interface with general education. *Journal for the Education of the Gifted, 20*(2), 138-154.

Sianesi, B. (2001, May). *Implementing propensiting score matching estimators with STATA.* Paper presented at the UK Stata Users Group, VII, London, England.

Smith-Maddox, R. & Wheelock, A. (1995). Untracking and students' futures. *Phi Delta Kappan*, *77*(3), 222.

Sparfeldt, J. R. (2007). Vocational interests of gifted adolescents. *Personality and Individual Differences, 42*(6), 1011-1021.

Stamm, M. (2007). Education of the gifted and social background. Findings to the hidden mechanisms of their interaction. *Journal of Sociology of Education and Socialization*, *27*(3), 227-242.

Stiefel, L., Schwartz, A. E., & Chellman, C. C. (2007). So Many Children Left Behind. *Educational Policy, 21*(3), 527-550.

Stiefel, L., Schwartz, A.E., Ellen, I. (2007). Disentangling the racial test score gap:

Probing the evidence in a large urban school district. *Journal of Policy Analysis and Management, 26*(1), 7-30.

Stiggins, R. (2003). Balanced assessment: The key to accountability and improved student learning. Washington, DC: National Education Association.

Teaching inequality: the problem of public school tracking. (1989). *Harvard Law Review*, *102*(6), 1318.

Tong, J. & Yewchuk, C. (1996). Self-concept and sex-role orientation in gifted high school students. *Gifted Child Quarterly,* 40, 15-23.

Tyson, K., Castellino, D. R., Darity, W. (2005). It's not "a black thing": Understanding the burden of acting white and other dilemmas of high achievement. *American Sociological Review*, 70, 582-605.

VanTassel-Baska, J. (2005). "Gifted programs and services: What are the nonnegotiables? *Theory Into Practice*, *44*,2, 90.

VanTassel-Baska, J., Feng, A. X. M., & Evans, B. L. (2007). Patterns of identification and performance among gifted students identified through performance tasks: A three-year analysis. *Gifted Child Quarterly, 51,*3, 218-231.

Watanabe, M. (2007). Tracking in the era of high stakes state accountability reform: Case Studies of Classroom instruction in north carolina. *Teachers College Record, 110*(3), 489-534.

Watanabe, M. (2008). Tracking in the era of high stakes state accountability reform: Case studies of classroom instruction in North Carolina. *Teachers College Record, 110*(3), 489-534.

Welner, K. & Burris, C. C. (2006). Alternative approaches to the politics of detracking. *Theory Into Practice*, 45(1), 90.