Within Classroom Peer Effects and Tracking: Assessing the Impact of Classroom Peer Variation and Ability Tracking with Disaggregated High School Data

A Dissertation
Presented to
The Academic Faculty

By

C. Kevin Fortner

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

Georgia Institute of Technology and Georgia State University

December, 2010

Within Classroom Peer Effects and Tracking: Assessing the Impact of Classroom Peer Variation and Ability Tracking with Disaggregated High School Data

Approved by:

Dr. Gary T. Henry
School of Public Policy
University of North Carolina – Chapel Hill

Dr. Robert J. Eger, III
Reubin O'D. Askew School of Public Administration and Policy
Florida State University

Dr. Monica Gaughan
College of Public Health
University of Georgia

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

Dr. David L. Sjoquist
Andrew Young School of Public Policy
Georgia State University

Date Approved: October 29, 2010

ACKNOWLEDGEMENTS

Thank you to all of the members of my dissertation committee.

I also would like to thank Robert for all of the helpful advice and time spent to move me forward in those initial years. Thank you to Gary for your feedback and advice as a chair, mentor, boss, and friend.

Thank you to the Andrew Young School of Policy Studies for providing additional resources through the Dean's Fellowship that allowed me to focus fully on my educational goals during my first three years in the program.

Thank you to Jennifer, John, Kyra, Jimmie, and Betty for your unconditional love and support. The list of things waiting to be done 'as soon as my dissertation is finished' can finally go in the other direction. Kyra, I look forward to the day soon when you are older than this project.

Thank you as well to Cathy Slade for asking questions that I couldn't answer, Dana Rickman for showing the way outside the classroom, and faculty at UNCG whose guidance in pursuing this path was invaluable.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This dissertation focuses on two distinct relationships: 1) classroom peer ability and student end-of-course test outcomes and 2) school tracking policy and student end-of-course test outcomes. Utilizing the education production function and hierarchical linear models, this dissertation contributes to the literature in the field of public policy by extending the work of previous scholars and focusing attention on these relationships in three high school subjects (English I, Algebra I, and Biology). In addition, I present a novel method for identifying tracking intensity within schools. The primary research questions addressed in the dissertation include: 1) To what extent does the ability level of classroom peers contribute to student test score performance?; 2) Does the variability of prior achievement within classrooms correlate with student test score outcomes?; and 3) Is there a relationship between school tracking policy and student test scores? Collectively these questions directly relate to policy options at the school, district, and state levels.

CHAPTER 1

INTRODUCTION

For over 40 years, researchers have struggled with inadequate data to understand how peer ability within schools or within classrooms impacts individual student achievement and attainment. Early studies, using relatively simple methods, focused on small-N experimental designs meant to reveal the differences in achievement between tracked and untracked classrooms (Slavin, 1990). Later efforts considered the racial distribution of students into schools and the effects that peer ethnic composition at the school level have on student achievement (Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld, and York, 1966). In recent years, more sophisticated methodological approaches in combination with the use of improved data brought increased attention to the topic, but failed to produce a consensus regarding the effect of peer ability on student achievement (Argys, Rees, and Brewer, 1996; Betts and Shkolnik, 2000a; Hanushek, Kain, Markman, and Rivkin, 2003; Zabel 2008).

The effects of peer ability on individual student achievement are especially important for policy purposes. Parental choices and public policy at a variety of jurisdictional levels impact how students select into or are assigned into schools and how students are assigned into classrooms within schools. Where assignment to public schooling is based exclusively on place of residence, parental decisions on where to live determine school selection. Public school policies impacting assignment may include alternative schools, open enrollment, magnet schools, concurrent enrollment, and charter school options (US Department of Education, 2004). Recent legislative efforts including the No Child Left Behind Act of 2001 (2001) and U.S. Supreme Court judicial decisions increased the prospect of school choice options within public schools (Zelman v. Simmons-Harris, 2002) and limited the ability of schools to assign students to schools based on race in order to increase diversity within schools (Parents Involved in Community Schools v. Seattle School District No. 1, 2007). All of these factors make the assignment of students to schools non-random occurrences. The non-random assignment of students into schools is one factor which makes it impossible to compare just the outcomes for students in different schools as a method for testing the effectiveness of differing school policies.

Once eligible to attend a particular school, sorting also takes place within schools based on student assignments to particular classes. Within schools, parental influence and school policies also impact student assignments into classrooms. Many motivations for the classroom assignment decision are possible. Sorting may take place on a random basis (as presumed in some studies). It may reflect the principal or school system's belief about matching the most effective teachers with students needing the most assistance. It may be heavily influenced by parental lobbying for specific classroom assignments, and sorting also occurs based on the timing of non-academic student choices. For example, students in a band class offered only during the day's first hour may tend to place students participating in band in the same classes later in the day since none of them attended English I during the first hour. Students pursuing a vocational track that occupies the second half of the school day must fill their academically focused coursework into a smaller number of possible classroom assignments. Principals can also assign students to classrooms based on beliefs about the efficiency or equity of various sorting arrangements including ability tracking or purposely untracked (heterogeneous ability) classrooms.

Formal ability tracking places students into classrooms based on the previous achievement level of students producing homogeneous ability groupings of students, which provides an instructor with an opportunity to tailor lessons to one particular level of student ability. Homogeneous grouping may yield efficiency gains due to the preparation of a more unified range of topics and pace for lessons. Tracking may also exist informally through intentional, but unstated, assignment to ability-grouped classrooms. An argument against ability grouping suggests that grouping students into classes based on ability may limit the chances for low-performing students to learn from higher-performing students (and for higher-performing students to make additional learning gains from assisting low-performing students). This counter argument creates tension between optimizing efficiency through maximization of student gains via tracked assignment to classrooms or mitigating gaps in learning between low and high achieving students via heterogeneous ability groupings. Intentional heterogeneous grouping or random assignment of students into classes provides more opportunities for interactions between students of different levels of ability or achievement than exist within the tracked environment.

If the influences of peers on student achievement are small, then policies that affect student sorting should garner little interest from policymakers. In this case, attention may focus on other policies that provide a more meaningful influence on student achievement. However, current evidence indicates that peer influences do matter (Argys, et al., 1996; Betts and Shkolnik, 2000a; Hanushek et al., 2003; Zabel, 2008;

Zimmer and Toma, 2000). For this dissertation, data was obtained and assembled that will improve on existing research in these ways:

1. This study will provide accurate estimates of the relationship between peer ability and student achievement by utilizing a micro-level dataset which identifies student placement into classrooms, connects those classrooms to individual teachers, and contains extensive control information on a variety of characteristics, including teachers.
2. This study will model the relationship between peer ability and student achievement for high school students, a currently understudied population in the peer effects literature.
3. This study will model the relationship between peer ability and student achievement in a science subject (Biology), a currently understudied area in education research, in addition to language arts and mathematics.
4. This study creates and utilizes a novel approach to determine the utilization of ability tracking within schools.

The primary research questions addressed in the dissertation are:

1. To what extent are high school students affected by the ability level of their peers in their classes?
2. How does the dispersion of student ability within classrooms influence student end-of-course test outcomes?
3. To what extent is student achievement in English I, Algebra I, and Biology affected by school tracking policies?

To provide unbiased estimates of the impact of peer ability levels on individual student achievement, this dissertation draws on the substantial use of the education production function in the existing education research literature as the framework for analysis. Investigation of these questions provides important information to address the efficiency and equity concerns of educators, researchers, policymakers, and parents.

The next chapter reviews the literature relevant to peer effects, tracking, and the education production function. Chapter 3 focuses on the research design, data, model specification, and research questions directly tested in the dissertation. Chapter 4 provides the results of the analysis, and chapter 5 focuses on the implications of the study for both policymakers and researchers.

CHAPTER 2

THEORY: PEER EFFECTS, TRACKING, AND THE EDUCATION PRODUCTION FUNCTION

Introduction

Chapter 2 focuses on the state of research literature on peer effects, tracking, and the education production function. Peer effects and tracking influence each other as policy decisions related to ability tracking must influence the classroom peers of students within tracked schools. Studies may attempt to identify only how one's peers influence achievement, but ability tracking as a policy uses peer configuration as a lever to change student outcomes. Studies vary their focus on the combination of these two concepts and their relationship with student outcomes.

The organization of students into classes expresses beliefs about what arrangements are best for student achievement and other school goals, such as parental satisfaction. If expectations are that random assignment to classrooms provides the best opportunities for students to make academic progress, we would expect to observe heterogeneous classes that reflect a school's overall student population. If assignment into classes differs from random assignment, it is a reflection of beliefs about what arrangements maximize student outcomes, welfare, etc. The era of accountability testing provides increased incentives for administrators to focus attention on student assignments that maximize student outcomes on statewide administered tests across subjects.

An opening section on the historical development of research into peers and tracking begins the chapter. The chapter continues with a discussion of more recent literature on peer effects and tracking from both the sociology research literature and the economics/policy research literature. Finally, research specific to the education production function is reviewed as a vehicle for examining the relationship between variables in education policy research.

Historical Development

Education research literature moved from a point where a school's contextual environment is considered to have an influence on student outcomes to models which more closely examined student peer groups and classroom assignments of students. This section identifies a few early articles that made major contributions to the study of peers and tracking in education. The Equality of Educational Opportunity[1] report (Coleman et al., 1966) broke new ground in education research by examining school-level peer effects and the context of educational achievement by focusing some of its research effort on the racial makeup of schools. In this view, peers influence educational achievement by changing the cultural environment of schools.

In an influential early research study, Summers and Wolfe (1977) pushed peer effects literature forward by utilizing individual level measures of student achievement and providing a comparison between analyses conducted with individual measures and school level aggregate outcomes. In addition, they introduce aggregate measures of grade level peers based on ethnicity and test score achievement as well as integrating a number of interaction terms to compare how school inputs might differ in effectiveness between different types of students. With regard to grade level peer effects, Summers and Wolfe (1977) find a positive effect on test score gains across the range of student scores and also

[1] Also known as The Coleman Report.

find that low-achieving students differentially benefit from the presence of higher percentages of high achieving peers.

By 1990, Slavin's meta-analysis of studies regarding ability-grouped peer effects in secondary schools concluded that empirical analyses of peer influence on educational achievement show no effect. Many of the works included in Slavin's meta-analysis do utilize experimental designs, but frequently rely on small samples and vary widely in terms of how tracking or ability grouping is defined and measured. Hallinan (1990) criticized Slavin's (1990) paper because it relies largely on comparisons between the average outcomes between tracked and untracked students. Hallinan points out the failure of studies using means (or averages) as outcomes to differentiate between potentially large variances in outcomes between high and low achieving students. Means may remain similar while within classroom variation has increased (or decreased). Hallinan also suggests that more contemporary and methodologically sophisticated studies to come will provide better data regarding the potential conflict between equity and efficiency inherent in tracking policy.

Hallinan (1994) later provides a defense of tracking explicitly stating that the purpose of tracking is to provide more efficient and effective instruction (p. 80). Hallinan's (1994) expressed concern is that the lack of an effect of tracking in research is due to poor implementation. She specifically points out five ways in which tracking is poorly implemented: placement into tracks based on non-academic considerations (such as course conflicts), variance between schools in track placement criteria, variance in track flexibility across schools, over representation of poor and minority students in lower tracks, and assignments to tracks based on social status. Hallinan acknowledged negative effects of tracking due to its tendency, in practice, to produce segregation, decreased social status for students in lower tracks, heterogeneous tracks, slow pace of achievement in lower track classes, and negative social / psychological consequences for students. Finally, Hallinan believes that if schools focus effort on mitigating the negative effects of tracking and correctly implement tracking, the efficiency benefits of tracking can increase overall student achievement in an equitable way.

Oakes (1994) responds to Hallinan (1994) by asserting that the negative effects of tracking are so intertwined with the policy that it is not possible to solve the deficiencies of tracking by implementing 'better tracking.' Oakes argues that problems with tracking implementation, such as heterogenous tracks identified by Hallinan (1994), are actually attempts by educators to address the inequities of tracking based on rigid track placements (p. 88).

By the early 1990s, research had still largely failed to provide enough information on the relationship between peers and student outcomes to effectively guide tracking policy. Despite the lack of consensus on effects, these studies introduced important concepts and definitions that would guide new research with disaggregated datasets and methodologically advanced approaches. The next section focuses on research in the areas of peer effects and tracking beginning with some examples from sociology and moving into relevant literature from the economics and policy field.

Peer Effects

This section begins with a look at three studies from the sociology field representing differences in the definition of peers in this field which contrast significantly with the economics and policy literature. It follows with an in depth focus on the research literature on peer effects and tracking in policy.

Studies in the fields of sociology and psychology commonly investigate relationships between peers and their impacts on school and behavioral outcomes of school age children by examining the mechanisms through which peers and tracking influence student achievement. It is common in these literatures for studies to define peers as personal relationships among individuals such as best friends or boyfriends and girlfriends; these studies also place more focus on individual survey measures of concepts such as academic engagement, motivation, and effort. This research provides a clearer picture of the variation of outcomes across different types of students within schools and even within classrooms compared to the focus in policy and economics on classroom, grade-level or school composition.

In his investigation of classroom tracking, Carbonaro (2005) examines the relationship between classroom track, student effort, and achievement. Carbonaro utilizes data from the National Education Longitudinal Survey of 1988. In this study, track identification can vary across subject, and teacher surveys ask teachers to identify classrooms as belonging to one of five categories: "honors or advanced, academic, general, vocational-technical-business, and other" (Carbonaro 2005, p. 34). Student effort is based on teacher survey information from two different time periods (and two different teachers) for the student in both 8th and 10th grade. A 10th grade math achievement test functions as the dependent variable in models examining achievement outcomes. Carbonaro (2005) compares students across schools and adjusts standard errors to account for clustering. He finds that students in higher tracks have higher academic achievement and exert more effort than students in other tracks, but that the returns to student effort are constant across tracks and differences in effort do not fully explain differences in achievement across tracks. All of the findings include controls for student characteristics including measures of gender, race, 8^{th} grade test achievement in four subjects, and socio-economic status (SES), but do not include controls for differences in teacher characteristics, school characteristics, or peer characteristics.

Cook, Deng, and Morgano (2007) define peers as friends that students identify by name to determine the influence of friendship groups on students' academic and social behaviors. Cook, Deng, and Morgano find that among the friend variables included in models the grade point average of friends (peers) is the most consistent and reliable peer characteristic predicting individual student achievement outcomes and also affects outcomes in the social behavior domain. While this study is not an example of novel findings within the sociological literature, in fact it supports many prior studies, and in addition, it improved on previous studies through its inclusion of a wider variety of control variables and measures of student behavior across domains including academic behaviors, social behaviors, and mental health. The Cook, Deng, and Morgano (2007) study provides a good example of the type of peer characteristics commonly examined within the current sociology literature.

While both of the two prior studies mentioned utilize ordinary least squares regression to identify the impact of peers on one's own academic or social outcomes, other studies have begun to use multi-level or hierarchical linear models (HLM). For example, Kiuru, Aunola, Vuori, and Nurmi (2007) investigate the association between a student's short and long-term education expectations and the student's peer group's educational expectations. Peers are defined as up to three friends within a student's classroom with whom these students "most liked to spend time" (p. 999). Kiuru et. al. find associations between peer group expectations and a student's own expectations,

Understanding at least some of the mechanisms through which peer effects may translate into changes in academic performance and behaviors, such as peer expectations, on student achievement may provide useful information to increase our understanding of the education process, but policymakers have more direct interest in policy changes that can be leveraged to improve academic outcomes. A focus on the relationship between classroom or grade level peers as an input in the education process and student test score outcomes that provides an estimate of the relationship between these variables could identify a policy lever to make classroom or school assignments more effective. If relationships between peer ability level and student test score outcomes are well documented, the mechanisms influencing the relationship may be able to be ignored for decision-making in the policy context. For example, understanding how friend relationships influence a student's test performance may be helpful when a parent or counselor has the capacity to influence friend bonds but is less useful as a mechanism for influencing student outcomes within a classroom or school. Schools can adjust classroom assignments to influence outcomes, but efforts to influence the friend relationships of students seem both more difficult and intrusive as a policy mechanism.

We turn now to the peer effects research which measures classroom or grade peers rather than social peers to examine relationships between peers and outcomes. Previous studies examine a variety of these relationships including how the average ability level of one's peers influences one's own test score outcomes, how the range or heterogeneity of ability within classrooms influences one's own test score outcomes, and how these relationships might vary across high or low achieving students. Studies also vary in the composition of peers and whether comparisons are made across schools or within schools.

Zimmer and Toma (2000) focus their research on the mathematics achievement of 13-14 year old students (typically 8th grade students in the US) utilizing data across countries. Peers in their model include all students within a classroom. This definition means that all students within a classroom have the same mean peer ability value. Data used in their study come from the International Association for the Evaluation of Educational Achievement and was collected in 1981. Zimmer and Toma (2000) present their empirical analysis as an attempt to resolve inconsistent evidence on the heterogeneity of peer effects between (1) Summers and Wolfe (1977), who find positive peer effects for both high and low ability students but greater increases for low ability students, (2) Henderson, Mieszkowski, and Sauvageau (1978), who find positive peer effects that are comparable across all types of students, and (3) Argys et al. (1996), who find that de-tracking schools will result in gains to low ability students that exceed the negative impacts on higher achieving students. Their findings include a positive impact of peer ability on all students with additional positive gains for low-achieving students in classrooms which are more diverse (Zimmer & Toma, 2000). A one point increase in the beginning of year test classroom mean is expected to increase student test scores by about 0.77 points (or about 0.15 standard deviation units) (Zimmer & Toma, p. 83). Zimmer and Toma produce strong evidence for positive peer effects for middle school mathematics students with a data set that includes a number of details on student and family characteristics but more limited data on teacher characteristics. In addition to analyses regarding the relationship between peer ability and student performance, Zimmer and Toma test for relationships between peer ability heterogeneity and student outcomes. Zimmer and Toma utilize the standard deviation of a student's beginning of year test scores as a measure of classroom heterogeneity and find that a one point increase in the standard deviation of student pre-test scores is associated with an expected increase in end-of-year student test scores of about 0.23 points (or about 0.028 standard deviation units) (p. 83). They do not extend these findings to assess the impact of ability tracking.

Hanushek et al. (2003) consider previous estimates of the relationship between peers and student outcomes as inconsistent. The authors characterize previous study estimates of peer effects as "open to widely varying interpretations" (Hanushek et al. (2003), p. 527). The results of investigations into the peer effects of ability tracking and racial composition are also viewed as mixed (Hanushek et al., p. 529). Hanushek et al. (2003) uses a cohort analysis that considers all students within a grade as peers. Utilizing mathematics test score data from public school students in Texas, Hanushek et al. find a positive relationship between peer achievement and achievement growth but no systematic relationship between variance in peer achievement and achievement growth. Their available data do not permit investigation of classroom tracking due to a lack of information on classroom assignments of students.

Work by Vigdor and Nechyba (2004) finds a positive relationship between 3^{rd} grade classroom peers and their 5^{th} grade test score outcomes in both reading and math for students in North Carolina. In addition, peer ability dispersion or classroom heterogeneity is included in some models, similar to the approach utilized by Zimmer & Toma (2000) and Zimmer (2003). This approach uses the standard deviation of student test scores in a previous time period as a measure of the dispersion in peer achievement. They find a positive relationship between classroom peer dispersion and test outcomes in math but not in reading (Vigdor & Nechyba, 2004). Finally, these authors also compared the results when peer definitions were grade level and classroom level and reveal substantial differences in results. Relationships are more likely to be observed when the unit of analysis for peers is the classroom and analysis conducted with covariates representing classroom and grade peers simultaneously show that classroom peers have substantially larger effects.

Hoxby and Weingarth (2006) utilize data from Wake County, North Carolina schools to estimate the effects of peers. Student performance data for students in grades 3 to 8 are used to estimate the impact of peers on student test performances using a combination of reading and math scores. Burke and Sass (2008) estimate the effects of classroom peers on student achievement in Florida public schools using data in math and reading separately for students in grades 3 to 10. For high school students (grades nine and ten only), Burke and Sass (2008) estimate a mean effect of a one standard deviation unit increase in peer ability as 0.0577 and 0.044 standard deviation units for math and reading respectively. Their estimation strategy uses both student and teacher fixed effects to estimate these peer effects. Zabel (2008) provides estimates of peer effects in his research on elementary grade students in New York City schools in both reading and math.

Table 2.1 Peer Effects Studies

	Grades; Location	Peer Definition	Estimated Impacts - Standardized Estimate of Peer Achievement
Zimmer & Toma 2000	Ages 13-14 (8th); International including USA	Classroom	0.15
Hanushek, Kain, Markman, & Rivkin 2003	3 – 6; Texas	Grade	0.27 to 0.43
Vigdor & Nechyba 2004	5; North Carolina	Classroom and Grade Comparisons	0.086 (Math Class OLS) 0.091 (Reading Class OLS)
Hoxby & Weingarth 2006	3 – 8; Wake County, North Carolina	Classroom	0.25 (linear in means model)
Burke & Sass 2008	3 – 10; Florida	Classroom	0.058 (HS Math) 0.044 (HS Reading)
Zabel 2008	3 – 4; New York City	Classroom	0.04 to 0.08

The previous studies view the effects of peers as distinct from tracking policy but hope to inform school policies by accurately estimating these effects. Table 2.1 above summarizes the peer effects focused literature discussed. Only one of these studies (Burke & Sass, 2008) provides estimates for high school student peer effects and these estimates are for achievement in general math and reading exams. None of the previously cited studies address peer effects in subjects other than grade level reading or math. The next section focuses on studies which address the subject of school tracking or grouping students based on their prior academic achievement.

Tracking

If peers exert an influence on student outcomes, there is likely to be a relationship between fully implemented tracking policy and student outcomes since tracking is designed to modify the classroom peers of students. The next set of literature is concerned primarily with tracking and to various extents integrates or ignores the direct influence of peers on estimates of the tracking effect. Policies might be evaluated without regard to how they work, but untangling this relationship between peers and tracking requires a consideration of both concepts and how they relate to student achievement.

Gamoran uses data from the High School and Beyond survey for the years 1980 and 1982 to determine how tracking affects student achievement in math and verbal exams (1992). Gamoran determines tracking status from sophomore student surveys in the 1980 year, while outcomes are based on test scores in the 1982 year. The author drops schools from the analysis where all respondents within a school reported enrollment in the same track (academic or non-academic). The study uses hierarchical linear models to test hypotheses but was limited to up to 36 student observations per school. The data set utilized lacks information on teacher characteristics and contains limited controls for student and school-level covariates, including peer measures (although overall school test score means are utilized). Gamoran finds a positive relationship between a student's individual level track status and test score outcomes two years later. Hoffer (1992) utilizes data from the Longitudinal Study of American Youth ending in the Fall of 1989 to compare outcomes for students in math and science courses between 7th and 9th grades. He also focuses on the between track outcomes within tracked schools. Track assignment is determined through a combination of student and teacher surveys that ask about track placement and school curriculum materials. Models include some socio-economic status variables for students. Hoffer finds that outcomes in grouped vs. non-grouped schools are about the same when controlling for prior test performance and a set of individual control variables, but that outcomes for low and high track students are negative and positive respectively.

Argys et al. (1996) investigate the impacts of ability grouping or tracking on student outcomes for tenth grade mathematics students. The authors look both at the likelihood of placement into different tracks and estimates of outcomes across different tracks. While this study provides adjustments for selection into specific tracks based on ability and other characteristics, the available data does not include direct measures of peer ability at the student or classroom level to adjust predicted track outcomes for the different ability level of peers in various tracks (teachers report the mean (or average) achievement level of students compared to their peers within the school) (Argys et al., 1996). The tracked status is identified in teacher surveys where instructors classify the classroom as composed of students "above average, average, below average, or widely differing (heterogeneous) achievement levels relative to other students in the school" (Argys et al., p. 628). The study concludes that lower ability students, those scoring in the bottom third of the test score distribution, would significantly benefit from assignment to heterogeneous versus low-ability tracked classes but that these gains would be fully offset by the losses of high-ability students moved from tracked classes to heterogeneously grouped classes. Argys et al. estimate that placing all students in heterogeneous classroom configurations would decrease overall achievement in mathematics by about 2 percent (p. 640).

Betts and Shkolnik (2000a) argue that previous attempts to identify the effects of tracking on student achievement are biased if students in schools which track students are compared to the average student in a non-tracking school. Studies prior to Betts and Shkolnik compared student outcomes in grouping (or tracking) schools to the average achievement in non-tracking schools. When unobserved student characteristics such as motivation and ability are correlated with track placement, biased coefficients result if the comparison group is average ability students in heterogeneously grouped classes. While Betts and Shkolnik find no difference in test score outcomes between students in schools identified as grouped or non-grouped by the principal using a variety of methodological approaches, other authors (Rees et al., 1999) have criticized their approach as one which compares outcomes in formally tracked (where a principal acknowledges tracking) versus informally tracked schools (where a principal denies tracking, but tracking implementation may still exist).

Figlio and Page (2002) changed the focus of tracking studies and use data from the National Education Longitudinal Study (NELS) to examine whether outcomes for low ability students are harmed by tracking. Instead of comparing the outcomes of low track students to high track students within the same school, the authors compare the outcomes for low performing students (those scoring in the bottom third of students based on 8^{th} grade math exams) to other low performing students in tracked vs. untracked schools. Their primary finding, that low performing students actually score higher when placed in tracked schools, are robust across a variety of methods for determining whether or not the school utilizes tracking (i.e. principal survey only; teacher survey only).

Later work by Zimmer (2003) assesses the impact of both tracking and peer effects on US public school mathematics achievement for 8^{th} grade students using a data set similar to that employed in earlier joint work by Zimmer and Toma (2000). Tracking is indicated on surveys of school characteristics within the data. He finds that tracked schools produce student gains that are 1.98 points higher than non-tracked schools (0.22 standard deviations). In addition, Zimmer finds that tracking enhances peer effects for low and average ability students when they are placed in classrooms with lower level peers. The effect of tracking on high ability students was insignificant. Zimmer uses a cutoff of students in the top and bottom 20 percent of the test score distribution to code high and low ability students respectively.

Two groups of authors have attempted to resolve questions of the relationship between ability tracking and student outcomes, but the relationship remains unresolved (Argys, et al., 1996; Betts & Shkolnik, 2000a; Betts & Shkolnik, 2000b; Rees, Brewer, and Argys, 1999). Betts and Shkolnik's (2000b) key critique of work by Argys, et al. (1996) and Hoffer (1992) is based on omission of an ability measure for a student's previous achievement potentially biasing estimates of tracking impacts. Betts and Shkolnik argue that if heterogeneously grouped classrooms are of a different overall ability level compared to tracked classrooms, biased estimates may result (2000b). Rees, et al. (1999) critique Betts and Shkolnik's (2000a) work by suggesting that the use of principal assessments on whether or not a school utilizes tracking make their work a comparison between formally and informally tracked schools, not a true comparison of actual school tracking policies.

Table 2.2 Tracking Studies

	Grades; Data	Track Definition	Subject	Results
Hoffer 1992	Grade 7-9; LSAY 1989	Teacher and school surveys	General science and math	Overall no effect, but shows differences in high track vs. low track gains
Argys, Rees, & Brewer 1996	Grades 8, 10; NELS 1988	Teacher survey	Math	Higher tracks have higher scores
Betts & Shkolnik 2000a	Grades 7-9,10-12; LSAY 1992	Principal survey	Math	No effect of tracking
Figlio & Page 2002	Grades 8, 10; NELS 1988	Teacher and principal survey	Math	Positive effects for low ability students
Zimmer 2003	Grade 8; SIMS	School survey	Math	Positive effect of tracked schools (0.22 SD)

The tracking literature has focused almost exclusively on tracking in the subject of mathematics with one study including the results of science tracking. Also, studies of tracking have focused on the effects in middle and high school grades with little attention being given to tracking practices in elementary grades. These focuses in the literature are most likely based on the perception that tracking is more commonly practiced in both higher grades and in mathematics. With one exception, all of the studies reviewed have focused efforts on identifying either peer effects or tracking effects in general mathematics or reading. Studies of these effects in specific high school courses such as algebra, English, and biology have not been published to date. Also, no studies examining the relationships between peer effects and tracking in biology have been published.

The definitions of peers and tracking vary substantially across these studies, but some preferred specifications have emerged. The prior economics/policy literature classifies peers as either classmates or schoolmates. Classmates are present during the delivery of lessons and may be an integral part of the learning process when instructors utilize class wide student-teacher interactions or student group interactions as a part of the learning process. Classroom level definitions of peers are more likely to yield statistically significant coefficients and fit better with our theoretical expectations regarding the mechanisms of peer relationships in educational settings. Studies have varied widely to the extent that they are able to fully identify the peers present in a student's classroom, even when they are able to accurately link a student and teacher.

Tracking is better specified as a school level variable versus the classroom level because even limited tracking for a small number of students modifies the remaining pool of students subject to assignment to other classrooms. A classroom survey of teachers may even fail to detect tracking when the instructor is asked to identify the academic ability level of a classroom as 'average' or not. Within a heavily tracked school, the average ability classroom might contain a group of students scoring at the average on a test of achievement from the prior year, but the range of student abilities within the classroom would be very narrow. In a school with no tracking and random assignment of students to classrooms, we would expect all instructors to characterize their classroom's ability level as 'average.' While average classrooms exist in both tracked and untracked schools, the range of test scores within each 'average' class would differ widely. Schools may vary significantly in how effectively they sort students based on prior test performance and how extensively they sort. The implementation of ability tracking may also vary across subjects. Schools may carefully sort students by ability in mathematics courses but pay less attention to ability grouping in social studies or literature courses. In order to determine the implementation of tracking within schools, an objective method for identifying tracking is needed. Identification of tracking within schools and specific subjects is further considered in Chapter 3. All of the policy oriented studies referenced in Table 2.1 and 2.2 above have used varying specifications of models known as the education production function. The next section focuses on changes to the education production function over time and the strengths and weaknesses of the approach.

Education Production Function

All of the recent policy-oriented studies reviewed use some form of the education production function to link educational inputs to student outcomes. This framework for analysis makes an explicit analogy between the production of educational outcomes to production in business where the product is some good or service. The education production function can also be thought of as an ecological model that recognizes the contributions of layers or contexts that may contribute to the success of an individual. Schools, classrooms, and parents are viewed as providing inputs that produce educational achievement as a product and attempt to maximize the production of achievement by adjusting the arrangement of relevant inputs into the most efficient combinations. In education, inputs typically include individual student and family characteristics as well as school and teacher characteristics. The education production function (EPF) provides a framework for unbiased estimation of the contribution of specific educational inputs to student outcomes.

In its fully specified form, the education production function requires data on characteristics of the student and their environment for all the years of their educational experience. For example, one's current achievement is a cumulative function with characteristics at the individual, family, classroom, and school level in each time period of schooling. For example, at the individual level, characteristics typically include a student's gender and ethnicity. Family characteristics might include the education level of one or more parents and the family's income level. Classroom characteristics typically include measures of the class size and contextual measures, such as the percentage of free lunch eligible students within a class. Additional measures of teacher characteristics may also be specified at the classroom level including experience or license qualifications. Finally, school level characteristics may include school size measures or contextual measures at the school level, such as the percentage of minority students attending a school. Many of these characteristics will vary over time as some individual, family, classroom, and school characteristics change from year to year.

Equation 2.1 presents the basic conceptual model of the education production function (Todd & Wolpin 2003, p. F15). The element T_{ija} represents an achievement measure for child i

Equation 2.1

$$T_{ija} = T_a\left[F_{ij}(a), S_{ij}(a), \mu_{ij0}, \varepsilon_{ija}\right]$$

in family j at age a. This specification recognizes that educational attainment is a cumulative function over time with age varying inputs that also vary in the relationship between each input and the age of individuals. F_{ij} and S_{ij} indicate the level of family and school resource inputs for a given student within a given family at each age, given by (a). A student's innate ability or "genetic endowment of mental capacity (determined at conception)" is represented by the term μ_{ij0} and does not vary over time (Todd & Wolpin 2003, p. F15). ε_{ija} represents measurement error specific to student i in family j in this specification and $T_a[\cdot]$ signifies that the relationship between inputs and the achievement measure are separate and may vary with age. Implementation of this model requires full information on an individual's family and school inputs at each age and a valid and reliable measure of the student's innate ability.

Given the lack of complete historical information on family and school inputs into a student's education, researchers have developed specifications of the education production function that require less data. Increasingly relaxed data requirements require increasingly strong assumptions to maintain validity. Todd and Wolpin provide a value added specification or value added model (VAM) for a "modified gain formulation with no endowment (where the lagged test score is a regressor)" as follows (2003, p. F19):

Equation 2.2

$$T_{ija} = X_{ija}\alpha + \gamma T_{ija-1} + \eta_{ija}$$

In Equation 2.2, T_{ija} represents the current year test score outcome for student i in household j at age a. $X_{ija}\alpha$ represents a combination of contemporaneous family and school resource inputs for the specific student. T_{ija-1} is the prior time period student achievement measure. And η_{ija} is an error term which encompasses both measurement error of the outcome score and a student's innate ability since it is unmeasured in the model. Todd and Wolpin warn that this specification of the EPF is sensitive to endogeneity bias because of the inclusion of a prior test score, if any relevant inputs are omitted from the model (p. F19). This endogeneity bias exists because both the prior test score measure and the current test score outcome are correlated with unmeasured mental capacity. The primary additional assumption required by this model compared to a model which includes each year separately is that the impact of inputs is constant for ages or grades for which outcomes are estimated in the model (p. F20). Models based on this specification are commonly referred to as value added models (VAM). VAM models may either utilize a difference between prior and current test score as the outcome variable (models which constrain the value γ in Equation 2.2 to one) or use a test score from a prior time period as an independent variable (Todd & Wolpin, 2003). Models using the difference of two test score outcomes as the outcome variable are termed gain score or gain specification models.

The gain specification requires slightly more restrictive assumptions compared to the specification of Equation 2.2 (Todd & Wolpin, 2003). In the gain specification, the return to prior achievement is fixed at a value of 1. This assumption becomes more difficult to accept when one considers the likelihood of measurement error in the two test administrations and that returns to prior performance might vary greatly based on where a student lies in the distribution of prior achievement. In addition, the gain specification is only appropriate where the two tests compared are closely aligned and comparable. This dissertation has two separate indicators of prior achievement, a student's 8^{th} grade test performance in both reading and math. These separate measures of prior achievement allow both scores to be integrated into the analysis reducing measurement error and providing an indication of student performance levels relevant to all three subjects in the study.

A specification test suggested by Todd and Wolpin (2003) is the inclusion of per pupil expenditure as a regressor. They suggest that if a model is fully and correctly specified through the inclusion of all relevant input variables that the influence of per pupil expenditure should be very near zero. Since expenditures influence specific inputs, such as class size or teacher quality, the influence of expenditures should be zero when all relevant inputs are present in a model.

Over time, studies using the education production function VAM specification have become increasingly popular (Hanushek, et al., 2003). Using this estimation technique places a premium on the inclusion of relevant control variables because of a known omitted regressor, the student's innate ability. In order to be considered valid, VAM studies must include other relevant inputs in the education production function and find achievement measures which minimize test measurement error.

Equation 2.3 modifies the prior equation to include a policy relevant variable, such as the effect of classroom peers on achievement. The terms $T_{ija}, X_{ija}\alpha$, γT_{ija-1}, and η_{ija} maintain their prior

Equation 2.3

$$T_{ija} = X_{ija}\alpha + Z_{ija}\beta + \gamma T_{ija-1} + \eta_{ija}$$

definitions, but the inclusion of item Z_{ija}, representing the prior test achievement of student peers within a classroom from a prior time period, allows the value of β to provide an estimate of the relationship between peer ability and a student's achievement measure in the current time period. The exact interpretation of β can vary depending on study design and the specification of the education production function (EPF) model.

Todd and Wolpin (2003) document the inconsistent relationships between some educational inputs and student outcomes in a variety of EPF studies. Todd and Wolpin theorize that these differences are primarily due to two factors: (1) improper comparisons between policy effect estimates and EPF *ceteris paribus* effects and (2) biased estimates of input relationships due to excessively strong assumptions in EPF studies (model misspecification). Randomized experiments provide an estimate of treatment effects due to policy changes and do not attempt to provide an estimate of the relationship between specific inputs and student outcomes, holding other variables constant. Todd and Wolpin theorize that families make resource decisions about their inputs into student achievement in response to the school's resource commitment to their child each year. If one wants to observe the estimated impact of a school policy change, no attempt to control for the level of family resource inputs is necessary or desirable. These estimates of policy impacts are, however, not comparable to an EPF parameter that provides an estimate of the impact of changing an input variable while holding all others constant. The authors argue that trying to compare experimental estimates to correlational or regression-based estimates of input variables will lead to confusion regarding the effects of policies since in reality these different types of studies answer different questions. In this dissertation, one goal is to provide unbiased or consistent EPF parameter estimates of the relationships between peer ability/peer dispersion and student test score outcomes for specific high school subjects. A further goal is to provide a policy relevant estimate of the relationship between student ability tracking and student outcomes. Providing policy relevant estimates requires the exclusion of control variables that may provide possible mechanisms for treatment effects. For example, if school tracking has a relationship with student outcomes and these outcomes operate through the manipulation of classroom peer ability, the inclusion of classroom peer ability as a regressor may understate the relationship between school tracking and student outcomes. When all regressors are included in models, a school tracking variable indicates the relationship between student outcomes and school tracking policy *net of classroom peer ability and all other included regressors*. This estimate may be useful as an EPF parameter but should not be considered an estimate of the possible policy impact of tracking.

The variables of interest in this analysis occur at the individual level (peer ability), classroom level (peer dispersion), and school level (tracking intensity). Ecological models theoretically recognize the contribution of varying contexts to the function of a system. In the school setting, students function within a classroom that lies within a larger school context. While we can theoretically recognize that similar students may perform differently when placed in a different school or classroom context, not all statistical models recognize this distinction. Hierarchical linear models explicitly recognize this arrangement of observations as nested and partition variance based upon the organization of observations. HLM models allow one to partition variance into the three levels discussed to estimate what proportion of variance in the outcome variable is between individuals within classrooms, between classrooms within schools, and between schools. This explicit recognition of the organization of observations avoids attribution of effects to levels which would be inappropriate if observations where assumed to be independently distributed.

In analyses of the relationship between peer ability and student end-of-course test outcomes, additional variables representing the 8^{th} grade test performance of a student's classroom peers are included in models and provide an estimate of the *ceteris paribus* effect of differences in a student's classroom peer group. These analyses recognize that each student within a classroom experiences a different set of classroom peers based on the prior ability level of all of the other students within a classroom, thus the variable exists at the individual level. Students in classrooms with high performing peers may make additional learning gains through exposure to higher performers in terms of group work, demonstrations of skills, or more intense discussions of class content. Students with low performing peers may similarly make more modest gains due to a lack of student to student learning within the classroom.

In models of the relationship between peer ability dispersion and student outcomes, the inclusion of an additional classroom level variable seeks an estimate of the relationship between student EOC test score outcomes and the diversity of student abilities within a classroom holding constant other model inputs. Prior research theorized that more diverse classroom environments might produce additional learning gains when opportunities exist for student to student learning and student ability varies widely (Argys, et. al., (1996). Narrow ability ranges may make the pacing of instruction more streamlined, but it is also possible that students retain more content when a variety of student skill levels are present within a classroom.

In the case of tracking, an additional school level variable indicates the level of tracking within a school. Initial models of this relationship drop variables which may confound the relationship between ability tracking and student outcomes, such as teacher characteristics, peer ability, and curriculum levels. Inclusion of variables which mediate the impact of student ability tracking would result in an attenuated coefficient on the ability tracking variable. The initial model of tracking in each subject attempts to measure a 'total effect' of ability tracking similar to a 'policy effect' or 'experimental estimate' in the language of Todd and Wolpin (2003). If a total effect estimate of tracking reveals a statistically significant relationship between a school's ability tracking policy and end-of-course test outcomes, additional models will attempt to see to what extent the relationship is mediated or 'explained' by additional variables such as peer ability or curriculum.

Despite the popularity of the education production function in the economics/policy literature, a number of critiques of its use exist. A primary criticism of the application of the EPF is that the production of education is fundamentally different from the factory model of inputs and outputs envisioned in the original production function. Critics argue that the provision of education may be so unique as to defy common findings that fail to account for differences in human interaction taking place within classrooms (Hodas, 1993). Like Monk (1992) and others who utilize the EPF, I counter that measurable characteristics of families, schools, classrooms, and individuals influence education production and the findings from well executed studies may inform policy and practice. Well-executed studies sufficiently operationalize theory and include relevant controls to aid in reducing omitted variable bias when a fully specified model is estimated. Omitted variable bias is a significant threat to the validity of model estimates.

The education production function provides a suitable framework for the investigation of educational inputs and policy effects and allows comparisons between studies focused on similar questions in various educational contexts which utilize the EPF framework. The data available for use in this study directly address a number of issues with omitted variable bias present in previous studies. While this chapter focused on the literature informing the study of this topic, Chapter 3 begins with a discussion of the design of the current study and the data and methods utilized to model the relationship between student outcomes and the variables of interest.

CHAPTER 3

DATA AND METHODS

Introduction

The two primary goals of this study are (1) to estimate the relationship between peer ability and student achievement and (2) to estimate the relationship between a school's tracking policy and student achievement in three high school subjects. Using the education production function as a theoretical framework to guide estimates of the relationships between peer ability, tracking, and student outcomes requires substantial information on students, teachers, schools, and student classroom assignments. Sufficient controls allow one to generate unbiased estimates of these relationships. With this goal in mind, chapter 3 begins with a discussion of the research design. It follows with detailed information on the model variables, model specification, and research questions addressed in the dissertation. Full descriptive data on the three data sets constructed for the analysis are also reviewed in this chapter.

Research Design

This section begins with a brief review of the areas of interest in this study and a description of the variables of interest used in the analysis. Next, it moves to a discussion of potential problems that the research design of the study is intended to resolve. Finally, a description of the relationship between these threats to validity and how the design of the study intends to address them closes the section. While the estimates of peer effects have varied substantially across studies, this study will focus on the relationship between peer ability and student outcomes for high school students enrolled in three subject specific end-of-course tested subjects (English I, Algebra I, and Biology). In addition to estimating the relationship between peer ability and individual student achievement in these understudied subjects, this dissertation will also assess the effects of ability tracking on the three subjects, compare their results, and assess to what extent potential mediators account for any observed relationships between student test performance and a school's tracking policies.

This study uses a quasi-experimental research design known as pretest-posttest nonequivalent groups design (Mark, Henry, and Julnes, 2000). This approach to determining the relationship between peer ability / tracking and student outcomes requires careful consideration of the ways in which the individuals exposed to differing peer groups / tracking policies may differ outside of the treatment received. For example, if very different types of students are found in tracked schools compared to untracked schools, simple comparisons of the outcomes in the two types of schools are invalid. Studies that fail to account for potential differences between schools, classrooms, and students are likely to generate biased estimates of the relationships between the variables of interest in statistical analyses.

Variables of Interest

Three sets of variables are of primary importance in this analysis: peer ability, peer dispersion, and tracking indicator variables. This section provides additional detail on these three sets of variables and how the measures are constructed.

I use two direct measures of peer ability in this study. First, a mean of peer ability based on classroom peers' standardized achievement on 8^{th} grade end-of-grade tests in reading or math (Peer Ability - Equation 3.1); second, a measure of peer ability dispersion within the classroom also based on classroom peers' standardized achievement on 8th grade end-of-grade tests in reading and math (Peer Dispersion - Equation 3.2). As discussed in the literature review of peer effects, these measures are consistent with previous literature on peer effects. The 'Variables of Interest' section concludes with a review of the tracking variable's creation.

Peer Ability

The peer experience for each student is unique since for a given classroom each student experiences a different set of peers, the entire class minus the student himself or herself. When considering the nature of peer learning, it is instructive to think about how the peer relationship is experienced individually for each person in a room. Taking the highest or lowest achieving individual out of a group can significantly alter the measured achievement level of the remaining individuals. The classroom experience for a single individual in a class might vary substantially depending on whether that individual is the top or bottom achieving student based on some common metric. Measurement of peer ability using the current outcome score is contaminated by the simultaneous determination of an outcome due to both characteristics of the individual student and the peer group. These two factors suggest that the measure of peer ability should be both individually determined and use measures from a prior time period.

Equation 3.1

$$\left(\overline{PA}_{(-i)cs}\right)_{(t-i)} = \left[\left((\overline{PA}_{cs})_{(t-1)} \times N_{cs}\right) - \left((A_i)_{(t-1)}\right)\right] \div (N_{cs} - 1)$$

Where $\left((\overline{PA}_{cs})_{(t-1)} \times N_{cs}\right)$ is the standardized mean peer achievement in grade 8 multiplied by the number of students in classroom *c* within school *s* and $\left((A_i)_{(t-1)}\right)$ is student *i*’s 8^{th} grade achievement score. This difference is divided by the number of students in the current classroom minus one to calculate a standardized mean of prior achievement for each student within the class.

The prior test score used for the peer ability measure in analyses varies by subject. For the series of models utilized to investigate research questions in English, I use the test performance of peers on 8^{th} grade reading tests to indicate peer ability. As expected, 8^{th} grade reading scores are more highly correlated with English I test scores than 8^{th} grade math scores, as assessed by Pearson’s correlation coefficient (0.796 for reading & 0.728 for math). In addition, prior performance on 8^{th} grade reading tests (coefficient of 0.527) is more predictive of English I test performance than prior performance on 8^{th} grade math tests (coefficient of 0.211) in the base model (Table 4.2 Model (1)). In Algebra I, eighth grade math scores are more highly correlated with EOC test scores than 8^{th} grade reading scores (0.732 for math & 0.597 for reading). In addition, prior performance on 8^{th} grade math tests (coefficient of 0.624) is more predictive of Algebra I test performance than prior performance on 8^{th} grade reading tests (coefficient of 0.103) in the base model (Table 4.3 Model (1)). Despite a higher Pearson’s correlation coefficient between Biology scores and 8th grade math scores (0.34 for reading & 0.722 for math), prior performance on 8th grade reading tests (coefficient of 0.429) in regression models including additional controls is much more predictive of Biology test performance than prior performance on 8th grade math tests (coefficient of 0.323) (Biology Results – Table 4.3 Model (1)). A student scoring one standard deviation higher than average on the 8th grade reading test predicts a Biology test score 0.429 standard deviations higher than average while an 8th grade math test score that is one standard deviation higher than average predicts a Biology test score only 0.323 standard deviations above average, holding other characteristics constant (Table 4.6 Model (1)).

Peer Dispersion

A classroom teacher faces a range of student abilities in each class that can vary widely by ability level. The measure of Peer Dispersion should be based on prior achievement, but is a factor that is a single measure indicating the environment faced by the classroom as a whole.

Equation 3.2

$$PD_{cs} = \sqrt{\frac{\sum_{i=1}^{n_{cs}}(X-\overline{X})^2}{N_{cs}}}$$

Where PD_{cs} is the standard deviation of peer achievement based on a student's 8th grade EOG test scores in reading or math depending on the subject. Similar to the peer ability measures above, I use 8th grade reading scores to indicate peer ability dispersion for analyses in English I and Biology, while 8th grade math scores serve as the prior ability measure for analyses in Algebra I. In order to address concerns that the number of students within a classroom would unduly influence a peer dispersion measure based on the standard deviation, I considered an alternative measure of the Peer Ability Dispersion measure. This measure is the interquartile range of standardized grade 8 EOG test scores within a classroom (difference between the 75th and 25th percentiles). For Algebra I, I calculated the interquartile range for classrooms and compared it to the Peer Ability Dispersion value. The two measures are highly correlated at 0.758.

Tracking

In order to test the effectiveness of tracked schools, it is necessary to establish a definition of school tracking. The dataset constructed for this dissertation provides the ability to use the assignments of students to classrooms to inform how extensively schools track students. This contrasts with prior work by Betts and Shkolnik (2000a) which use principal surveys to define tracked schools, and Argys et al. (1996) which use teacher survey responses to determine the tracked status of classrooms. The extent to which schools engage in ability tracking may vary across schools and by subject. The approach within considers each subject separately to uniquely identify whether or not a school engages in ability tracking in a subject, but also how precisely schools assign students on the basis of similar ability. The variations are referred to as differences in ability tracking intensity among school tracking policies. Schools which have the least variation in prior peer ability within classrooms are considered to have the highest levels of tracking intensity. By using an approach based on the amount of actual ability grouping within classrooms, I am able to avoid relying on self-reports of principals or teachers that may be subject to bias from social desirability or differences that occur by subject or due to chance.

The process to determine school tracking begins by simulating random assignment of students into classes ten times. I then calculate the difference between the average classroom standard deviation based on 8th grade EOG test scores of the ten random assignments to the actual average classroom standard deviation of 8th grade EOG test scores to determine the difference in mean standard deviations by school.

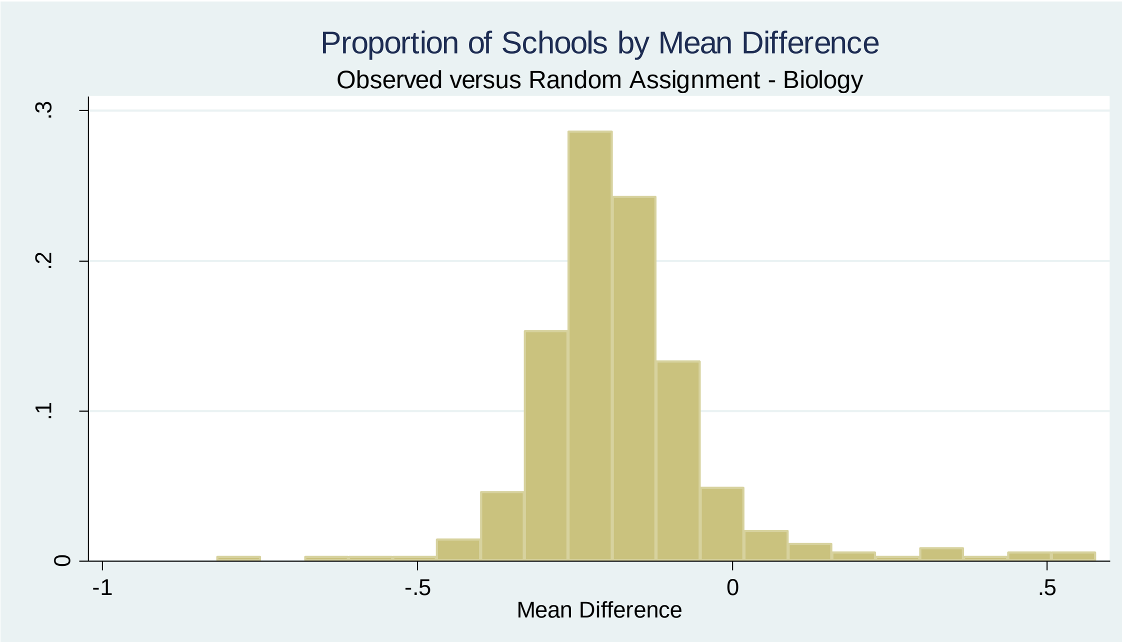

Figure 3.1 – Proportion of Schools by Mean Difference - Biology

Figure 3.1 displays the distribution of schools based on the difference in average classroom standard deviation between the average of random assignments and observed assignments in Biology. Schools further to the left represent classrooms with less variation in prior ability than expected compared to random assignment. Differences near zero and higher indicate schools which appear to assign students in a manner that approximates random assignment based on prior 8th grade EOG achievement. Schools are then split into quartiles based on the mean difference value.

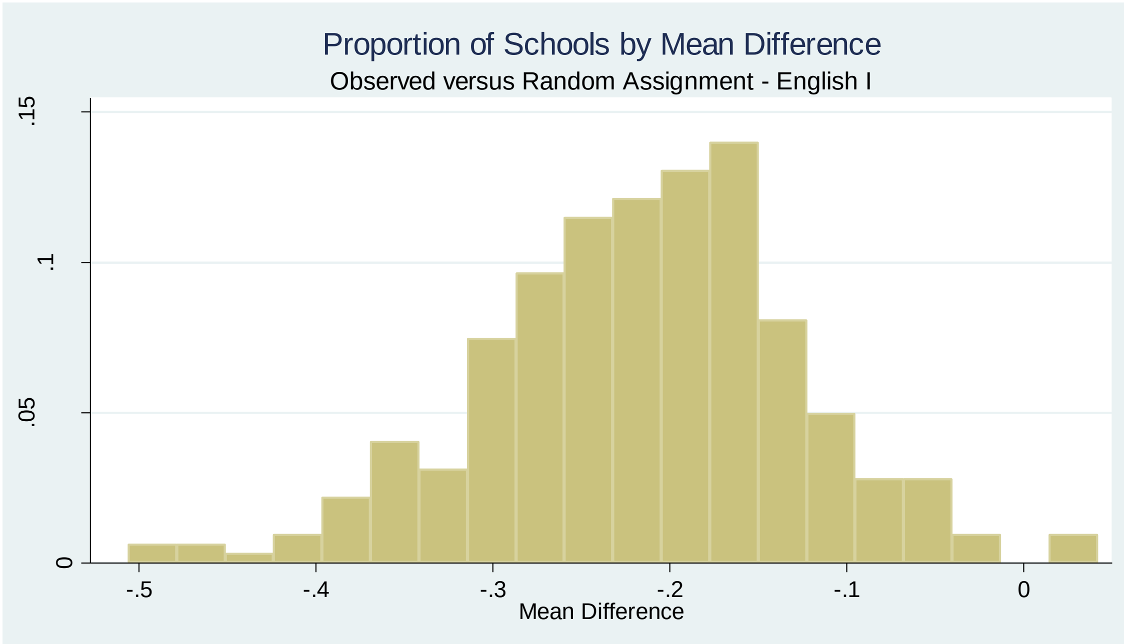

Figure 3.2 – Proportion of Schools by Mean Difference – English I

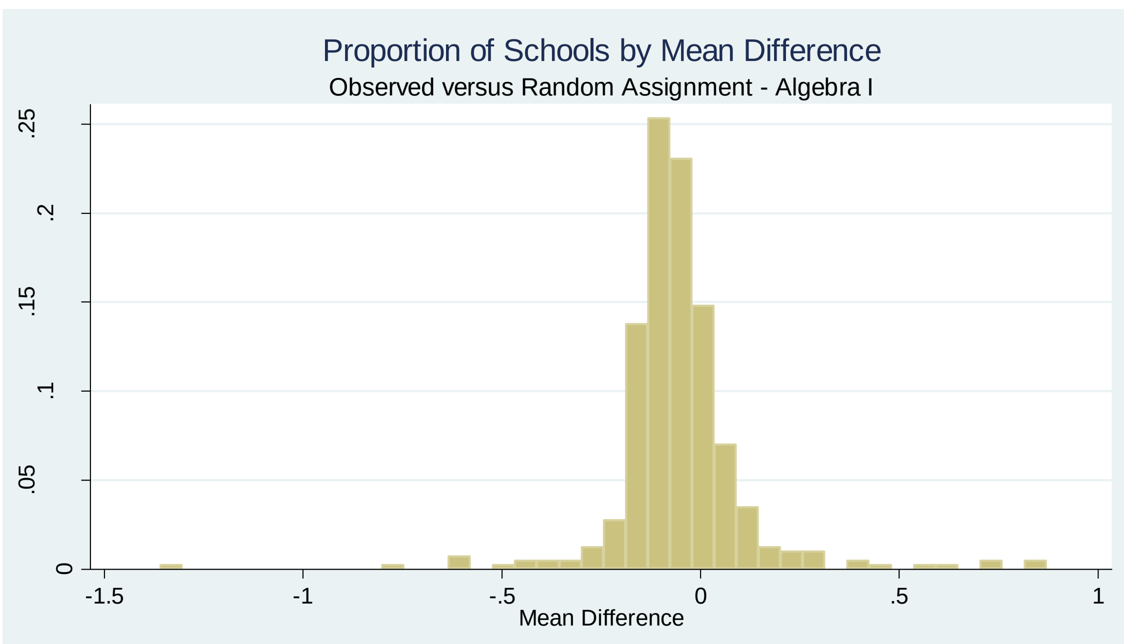

Figure 3.3 – Proportion of Schools by Mean Difference – Algebra I

Two additional figures (Figure 3.2 and Figure 3.3) display the distributions of the mean difference values for English I and Algebra I respectively. In English I, few schools appear to organize classrooms in a way that would be described as no tracking schools (values above zero). A fair proportion of schools though do have mean difference values that are less than -0.15 and the distribution of tracking intensity is more evenly dispersed for students enrolled in English I. It is possible that administrators consider a wider variety of factors that just prior test performance in making classroom assignment decisions for English I students. The distribution of mean difference values for schools based on Algebra I classrooms reflects an overall less intense tracking policy compared to Biology and English I. This is partially the result of less variation in prior test score performance as advanced students who enrolled in Algebra I in 8th grade are excluded if they move on to another math course in high school. These differences in distribution are reflected in the comparison between quartiles discussed below.

Table 3.1 – Determining Tracking Intensity

English I			
	Schools[2]	Mean	Std. Dev.
Tracking Intensity			
High	81	-0.33	0.05
Moderate	80	-0.24	0.02
Low	81	-0.18	0.02
Negligible Ability Tracking	80	-0.11	0.05
Algebra I			
	Schools	Mean	Std. Dev.
Tracking Intensity			
High	100	-0.21	0.17
Moderate	100	-0.09	0.01
Low	100	-0.04	0.02
No Ability Tracking	99	0.13	0.18
Biology			
	Schools	Mean	Std. Dev.
Tracking Intensity			
High	87	-0.30	0.07
Moderate	86	-0.22	0.02
Low	87	-0.16	0.02
No Ability Tracking	86	0.01	0.20

Table 3.1 displays the decreasing intensity of tracking as one moves from schools ranking from high to none. I use these ranks to create indicator variables for high, moderate, and low intensity tracking respectively. The lowest rank schools in terms of difference values are on average equivalent to un-tracked schools and will serve as the

[2] Tracking intensity values are calculated for all schools whether or not they are included in the final analysis due to data limitations. Approximate ¼ of students are present in each category in the analysis data sets. Grade level variations and subject offerings result in different school counts for each subject. School counts vary due to the grade level configurations of schools and school course offerings.

reference group in analyses related to tracking intensity. A one-way analysis of variance test confirms that the variance across the four groups of schools exceeds the within groups variance at statistically significant levels (p-value less than 0.000).

Tracking intensity does vary by subject. Even the lowest level of ability tracking for English I students represents an average difference value of negative 0.11. A greater number of schools appear to engage in ability tracking in English I compared to Algebra I and Biology. Biology appears to mirror English I values for the high, moderate, and low Tracking intensity difference values, but the No Ability Tracking category is very near zero for Biology. The compression of ability levels is evident in Algebra I as the difference values show that the highest levels of tracking in Algebra I are approximately equivalent to the moderate levels of tracking in English I and Biology. Also, the gap between the High and Moderate levels of tracking for Algebra I is higher than the gap between levels in Biology and English I. Should the results vary by subject in analyses of tracking effects, these differences may be useful in explaining divergent results. The research design section of the chapter continues with a focus on the potential problems which threaten the validity of findings when using the education production function to estimate these relationships.

Researchers encounter several problems when attempting to determine the relationship between classroom peers, tracking, and individual level student outcomes. The most serious threat to the estimate of the relationship is omitted variable bias. If variables related to student value added performance are omitted from models and those omitted variables are also correlated with peer and tracking variables, the coefficients produced in statistical analyses are biased. This might results if students with greater academic ability are differentially placed in classrooms with higher ability peers or higher performing instructors. If relevant measures of student ability or teacher effectiveness are excluded from models, this type of omitted variable bias would impact model estimates. Additional threats to the validity of findings are due to treatment and selection bias. If tracking policies are implemented in schools where tracking is likely to be effective and not used in schools where principals expect it to be ineffective, statistical analyses generate biased estimates that confound the true effects of tracking (if any) and the effect of treatment bias. Additional control variables capture differences between schools that influence whether or not a school pursues a tracking policy (i.e. school size or rural/urban designation). Teacher control variables are critical to rule out student gains that are due to true peer or tracking effects versus classroom assignments that match more effective teachers to higher (or lower) performing groups of students. Selection bias occurs when enrollment into treatment is non-random. It is referred to as self-selection bias when the individual chooses a treatment herself. For example, a comparison of treatments for quitting smoking might be biased if individuals more likely to benefit from group therapy select group therapy from among a selection of treatment options. This could be the case for Algebra I, because substantial numbers of students enroll in Algebra I during their 8^{th} grade year and are excluded from the sample based on their course taking decisions. This course taking pattern substantially alters the sample of individuals included in the Algebra I high school analysis compared to English and Biology I and could generate biased estimates if decisions to enroll in Algebra I in high school are correlated with peer quality or tracking expectations. The fourth significant threat to studies of this type is sampling or non-response bias. This type of bias also generates biased coefficients that result in estimates that lack external validity or the inability to extend the results to a more general population. This type of bias is typically due to patterns of non-response that leave a substantially different population of individuals for analysis while persons of certain characteristics are excluded. Analyses on a biased sample of individuals cannot be extended to expectations for the overall population of individuals.

The analysis is designed to minimize the types of bias described above. Addressing omitted variable bias is addressed in a straight forward way only if there is wide agreement about what factors influence the dependent variable under investigation, student test score outcomes in this case. Equation 3.3 presents more detailed but still reduced form specification of the education production function discussed in Chapter 2:

Equation 3.3

$$Y_{ics} = \beta_0 + Z_1\left(\overline{PA}_{(-i)cs}\right)_{(t-1)} + \beta_1 A_{ics(t-1)} + \beta_2 X_{ics} + \beta_3 F_{ics} + \beta_4 T_{cs} + \beta_5 C_{cs} + \beta_6 S_s + \varepsilon_{ics}$$

where

Y_{ics} is the EOC test outcome in time *t* for student *i* in classroom *c* in school *s*;

Z_1 is the estimated relationship between peer ability and the EOC test outcome;

$\left(\overline{PA}_{(-i)cs}\right)_{(t-1)}$ is the peer ability measure for student *i* in classroom *c* in school *s*;

$A_{ics(t-1)}$ is 8th grade achievement for student *i* in classroom *c* in school *s*;

X_{ics} is a vector of individual inputs for student *i* in classroom *c* in school *s*;

F_{ics} is a vector of family resource inputs for student *i* in classroom *c* in school *s*;

T_{cs} is a vector of teacher characteristics in classroom *c* in school *s*;

C_{cs} is a vector of classroom characteristics in classroom *c* in school *s*;

S_s is a vector of school characteristics common to all students in school *s*;

$\beta_1 - \beta_6$ estimate the relationships between controls and the EOC test outcome; and ε_{ics} is an error term specific to student *i* in classroom *c* in school *s*.

The term $\left(\overline{PA}_{(-i)cs}\right)_{(t-1)}$ specifies the peer ability level of an individual student's classmates. In many of the studies conducted in this area, researchers lack complete information on the classroom peers present for an individual student. For example, in the studies (Argys, et al., 1996; Figlio & Page, 2002) utilizing data from the NELS 1988 data set, up to approximately 24 students were selected for sampling in each school ("Quick Guide," 2002). Similarly, studies (Betts & Shkolnik, 2000a; Hoffer, 1992) using the LSAY dataset also do not have complete information on classroom peers or their ability levels (Betts & Shkolnik, 200b, p. 25). Most of the studies focused on peer effects (Burke & Sass, 2008; Hoxby & Weingarth, 2006; Vigdor & Nechyba, 2004; Zimmer & Toma, 2000; Zabel, 2008) do include information on the characteristics of classroom peers but are typically limited in their ability to measure extended classroom or teacher characteristics including curriculum differences and observable teacher quality measures that might impact classroom achievement. In the present study, access was provided to the full population of student rosters which allows precise definition of peers within classrooms, linking of classrooms to teachers including many measured characteristics of teacher quality, and information on the curriculum level of classes.

The term $A_{ics(t-1)}$ is typically represented by a single measure of math performance from the prior school year (in studies where math achievement is the outcome variable of interest). Alternatively, prior performance measures may include a pretest of the same type in the same school year, such as the math exams administered by the International Association for the Evaluation of Educational Achievement used in the Zimmer and Toma (2000) study. This provides a pre-test value and a second post-test assessment that serves as the outcome variable (Zimmer & Toma, 2000; Zimmer, 2003). Other times, the performance measure is more than one year prior, such as the studies using NELS data (Argys, et al., 1996; Figlio & Page, 2002). In this study, students may be taking any of the three subjects studied (Algebra I, English I, or Biology) in any of the high school grades 9-12, but have no pretest scores specific to these subjects. Students do take standardized curriculum referenced tests at the end of grade 8 in both reading and math, and both of these subjects will serve as prior achievement measures for students in a previous time period. At the time of data collection, the state of North Carolina did not conduct any statewide testing in science achievement for students enrolled in elementary or middle grades. Therefore, there is no precise measure of a student's Biology achievement in a prior time period, but the 8th grade math and reading exams are intended to serve as measures of a student's ability to be successful in the Biology course. The measures section provides more detail on the variables utilized to account for prior student achievement.

The term X_{ics} includes terms representing a variety of individual characteristics and family resources that may influence individual achievement. Individual student characteristics include gender and five ethnicity indicators based on the ethnicity categories defined by North Carolina's Department of Public Instruction. Two additional variables indicate students who are underage and overage. These variables compare a student's date of birth to the cohort cutoff age to identify students that are too young or too old for their grade level. The underage variable indicates students who may have skipped grades or started school early due to advanced abilities while the overage variable provides an indicator for students starting school later than expected or those who were retained in a prior time period. Two additional variables note students with disabilities and students identified as academically or intellectually gifted. Students in these two categories may learn at decreased or increased rates compared to students without exceptionalities, the reference group. Excluding Hoffer (1992), which included a measure of IQ in analyses, none of the other cited studies include any variables that would indicate differences between students in rates of learning and require the assumption that all students' gains come at approximately the same rate. Burke and Sass (2008) use a student fixed effects approach that allows this restriction on learning rate assumptions to be lifted, but still requires the assumption that student learning rates are fixed across time.

Students learning the English language may also learn at different rates compared to native English speakers. Two indicator variables identify students that previously received Limited English Proficiency services and those receiving those services in the current school year. Three variables also indicate a student's grade level with ninth grade students serving as the reference group in models. Finally, a student's number of absences is included in models to differentiate students based on the number of missed school days during a year.

Family resource variables are indicated by F_{ics}. Variables signifying a student's status as free or reduced lunch eligible differentiate students based on family income. The remaining family resource variables indicate the parental education level of students. Parental education level is divided into four categories: less than high school graduate, high school graduate only, some college, and a bachelors degree or higher. A parental education level of high school graduate only serves as the reference group in analyses. Table 3.3 summarizes the individual level control variables included in models and provides a rich set of covariates intended to remove the influence of these factors on estimates of the coefficients of interest.

Teacher level controls represented by T_{cs} include measures of attainment of a master's or higher degree, licensure status, competitiveness of the teacher's undergraduate institution, the teacher's experience level, how the teacher performed on assessments of general academic ability and teaching knowledge, and additional salary payments not directly tied to levels of experience or qualifications. Teachers holding master's or higher graduate degrees are coded as having an advanced degree. Three variables related to the types of teacher licenses issued in North Carolina are included. First, a variable indicating 'Other Licensure' indicates teachers that do not hold the standard initial or continuing license issued by the state. These include emergency, lateral, or provisional license types. A second indicator variable notes teachers holding a Nationally Board Certified credential issued by the National Board for Professional Teaching Standards. The third variable indicating infield status denotes teachers whose license area matches with the subject being taught. For Algebra I and English I, the license areas are Secondary Math and Secondary English respectively. Teachers are coded as infield for Biology I if they hold a license in Secondary Science or Secondary Biology.

The competitiveness of the of the teacher's undergraduate degree is indicated by the Barron's rating of the institution. Teachers with degrees from institutions rated as most or highly competitive are indicated with the Barron's rating variable. The years of experience values are broken into a series of categorical dichotomous variables with the category of five to nine years of experience serving as the reference group. Teacher test performance is standardized by test type and then averaged for each individual teacher. The teacher test performance is a continuous variable that represents the teacher's test score performance relative to all teachers taking the same test whether it is an SAT, GRE, specific PRAXIS, or NTE exam. The remaining teacher variable indicates the level of non-certified teacher compensation. This is the compensation amount paid to the teacher from local supplements and bonus pay codes and is not directly tied to the teacher's years of experience, NBC, and advanced degree status. Many of these measures are commonly unavailable in studies of this type. In addition, many studies are unable to precisely link students to their teachers. Among the peer effects studies reviewed, only Zimmer and Toma (2000) explicitly model individual level characteristics of teachers. Burke and Sass (2008) utilize teacher fixed effects which are intended to control for these teacher differences, and Hanushek, et al. (2003) used grade level averages of teacher characteristics in their models. This analysis includes a number of variables to adjust outcomes based on the characteristics of the classroom teacher.

The term C_{cs} represents a series of variables identifying the characteristics of classrooms, excluding teacher and peer achievement characteristics. These characteristics include the class size, the curriculum level of the class (advanced or remedial), and variables representing the contextual characteristics of the class such as income and ethnicity percentages of students. Finally, a number of characteristics indicating observable characteristics of schools is designated by the term S_s. School characteristics include total per pupil expenditures, school size, teacher turnover, percentages of students by income and ethnicity categories, and rural/urban designations of the school's location. All of these control variable terms are more fully discussed in the measures section of this chapter.

The present study includes a variety of variables not typically included in studies investigating relationships between peer ability, tracking, and student outcomes. This study includes new variables at the individual, teacher, classroom, and school levels. In some studies, these variables may be excluded due to the use of fixed effects modeling, but none of the cited studies include the wide variety of covariates included here. This section continues with a consideration of possible bias due to treatment and non-response bias.

For treatment bias to influence the results of this study, tracking would have to be selectively implemented in settings or with students and teachers where it is likely to be effective and not implemented in settings where it is likely to be ineffective. Similar to omitted variable bias, this bias requires that the factors of selecting schools for treatment be related to tracking and student performance but unrelated to the control variables included in models. If tracking is only effective and practiced in schools with large enrollments in suburban settings, excluding these control variables would generate positive tracking coefficients. In order for remaining treatment bias to exist in model estimates, the factors influencing the implementation of tracking must be both correlated with student outcomes and un-correlated with the existing model controls. Similarly, estimates of peer effects can be biased upward if higher achieving students are systematically assigned to more effective teachers and relevant teacher characteristic controls are excluded from models. This treatment bias could exist in this analysis only if teachers with certain unmeasured characteristics are differentially assigned higher achieving classroom groups. The extensive controls include many characteristics of schools, classrooms, teachers, and students to effectively limit the possibility that treatment bias alters the study results.

Another type of bias that may influence findings is self-selection bias. When individuals have the opportunity to choose whether or not to enroll in certain courses or treatments, determining whether an observed effect would be valid for other persons choosing to receive the same treatment becomes extremely difficult. In the case of English I and Biology, very large numbers of high school students enroll in these courses in their high school years. In Algebra I, however, about 1/3 of students take this course during their final year in middle school, grade 8. If students were randomly selected to take Algebra I in middle school, we could ignore this selection issue and estimate relationships on the observed patterns of data in high school only and presume that any relationships would hold for students had they waited to take Algebra I in high school. In actuality, students with the most advanced skills are the ones that are most likely to take Algebra I in 8^{th} grade. Since the goal in this analysis is to estimate relationships that rely on the characteristics of classroom peers, these selection patterns are a substantial threat to the validity of the findings if the influence of this selection by students into Algebra I in 8^{th} grade is not addressed. The research design addresses this selection through a weighting approach that over weights students who have similar observables to those students typically taking Algebra I in 8^{th} grade and under weights students who, given their observable characteristics, would be very unlikely to enroll in Algebra I in grade 8.

This weighting approach, known as Heckman correction or inverse mills ratio weighting is an effective method for addressing selection bias (Wooldridge, 2003). Additional information on the weights utilized is found in Appendix A.

The combination of more complete information on students through a population sample and the inclusion of additional variables decreases the concerns related to omitted variable bias and treatment bias in the present study. Inverse mills ratio weights address concerns that Algebra I estimates may be suspect due to self-selection bias. The next section considers the impact of non-response bias on the study design.

In studies based on survey data, patterns of non-response may lead to substantial differences between the intended population under study and the sample of individuals who actually respond to the survey. This study utilizes administrative data on all individuals enrolled in North Carolina public schools and teachers in these schools. Missing data does occur in this study due to failures in matching of teachers to classrooms and in cases where individual students cannot be matched to their test scores from prior years. Matching non-response decreases the sample of cases substantially in this study. The Data section of this chapter begins by considering the non-response patterns present in this analysis.

Prior research has considered data elements that would improve the ability of researchers to correctly identify relationships between tracking and student achievement. The specification of the education production function used in this study contains many of the elements included in the “optimal data set” for identifying the effects of tracking on student achievement discussed by Betts and Shkolnik (2000b, 25).[3] Table 3.2 summarizes the elements identified by Betts and Shkolnik (2000b). The two elements

Table 3.2 - Optimal Data Set for Estimation of Tracking Effects (Adapted from Betts and Shkolnik, 2000b)

Data Characteristic	Present in Data Set
Within classroom mean achievement	Yes
Within classroom achievement dispersion	Yes
Curriculum differences across classrooms	Yes
Within classroom grouping variables	No
Within classroom curriculum differences	No

missing from this specification include within classroom grouping (assigned students within a class to subgroups to work on assignments or perform tasks) and possible within classroom curriculum differences (where an instructor might differentiate instruction across the classroom). These omitted factors would bias estimates if there were systematic variation across observations that were correlated with other independent variables included in the model. This study's specification significantly improves on existing research regarding the relationship between peer ability/tracking and individual student achievement by integrating measures of mean peer achievement, peer ability dispersion, matching individual students to their classroom teachers, and providing measures of curriculum differences across classrooms.

The research design of this study addresses the most serious threats to extending the findings in this study to other locales through a combination of improved data availability, modeling approach, and weighting adjustments. The remaining sections of

[3] See Table 2.1 for a listing of these data elements (adapted from Betts and Shkolnik, 2000a, 25).

this chapter address the specific measures used in models, the precise model specification, the data, and the specific research questions used in this study.

Student Outcome Measures

In this study, the outcome variable is a value-added measure of student performance on End-of-Course (EOC) tests by high school students in North Carolina. The EOC subjects included in the analysis are English I, Algebra I, and Biology. EOC scores are standardized (z-scores) within subject across all students taking a given test, regardless of their grade level.

Individual Student Characteristics

Within the education production function framework, it is necessary to include individual level variables that reflect the personal characteristics of students and the resource inputs available to students either from themselves or their family. Prior research supports including controls or input measures for items such as prior test performance, family income, student race/ethnicity, English language learners, special education enrollment, gender, and parental education levels (Figlio and Page, 2002; Zabel, 2008). While not all variables represent input characteristics (i.e. ethnicity), these variables do correlate with student performance, and their exclusion could introduce omitted variable bias into the model estimates. Value added models typically include a range of individual student characteristics to control for differences in student outcomes. This section introduces the individual level variables included in models.

The primary control variable key in the value added model specification discussed is a student's performance on 8^{th} grade end-of-grade tests in both reading and math.

These scores are standardized by year for all student taking tests within a given year and are designed to adjust for historical inputs into a student's educational achievement following Hanushek, et. al. (2003), Hoxby and Weingarth (2006), Vigdor and Nechyba (2004), Zabel (2008), and Zimmer and Toma (2000). No prior scores focused specifically on achievement in the subjects under study (English I, Algebra I, or Biology) are available within this data set. In addition to the standardized grade 8 reading and math performance variables, squared and cubed versions of these variables are included in models to incorporate any non-linear relationships between prior achievement in reading or math and the subject under study. These interaction terms allow the relationship expressed between variables to be expressed as curves instead of only as a linear function where a unit change in the independent variable results in a constant expected increase (or decrease) in the outcome variable over the range of values.

A student's innate ability is not frequently included in education production function models. When available, it is often operationalized as an IQ test or similar measure of ability thought to be a reasonable measure of an individual's ability to learn independent of the specific instrument utilized as an achievement measure. Omission of an additional student ability measure is expected to generate upward bias of coefficients on individual characteristics correlated with a student's ability including family background measures and income (Hanushek, 1997; Zimmer and Toma, 2000). This slight upward bias could exist due to differences in the rate of learning across students. Two dichotomous variables in this study relate to the ability levels of students. The North Carolina Department of Public Instruction compiles data on student exceptionalities (characteristics of students that make them eligible for differentiated instruction programs). In models, one variable indicates that a student is coded as being academically or intellectually gifted. A second variable indicates students with disabilities. While prior test score performance would account for differences in the achievement levels of students at the end of grade 8, these additional variables control for differences in learning rates between these two groups of students and other students who neither have disabilities nor are gifted.

Indicator variables designate students who are over/under age for their cohort and student grade level as these classes serve students in various grades. No data is included that allows a determination of whether or not a student was ever retained. Using the student's date of birth and the enrollment cutoff in North Carolina public schools it is possible to construct indicator variable for students who are too old or too young for their cohort of enrollment. These variables are represented as under age and over age indicator variables. Also, the grade level of students is represented with dichotomous variables for each high school grade, nine through 12. In some models, a student's status as a 'high' or 'low' achieving student based on test scores in grade 8 (top or bottom 20% of students in their cohort in either reading or math) allow for specific testing of hypotheses related to high and low ability students. This cutoff for high and low ability students is the same as that used by Zimmer and Toma (2000).

Additional independent variables at the student level include student gender and the student's race/ethnicity (North Carolina Department of Public Instruction designated categories of American Indian, Black, Hispanic, White, Asian, and Multi-racial). Variables on family income and education level relate to the resources available to students and their families within the EPF framework. Free or reduced lunch status or family income takes on dichotomized values on variables for free lunch eligible, reduced lunch eligible, or eligibility status missing. An additional variable representing parental education is recoded into education categories representing less than high school graduate, high school graduate, some college, or baccalaureate or higher degree. Two additional variables indicate a student's status as a current or prior recipient of Limited English Proficiency services (LEP student status). A continuous student level variable indicates the number of absences during the school year.

Teacher Characteristics

The education production function requires measures of teacher inputs which are correlated with contributions to student's test score outcomes. These measures are frequently described as indicators of teacher quality. Common measures include years of experience, attainment of a master's degree, National Board Certification status, teacher test scores, undergraduate educational institution quality, and teacher salary levels. The degree to which these commonly used measures are related to test score outcomes varies. Despite studies failing to show positive impacts of attaining a master's degree on student outcomes (Hanushek, 2003a), this and similar controls exist in models to provide further evidence regarding these findings and to address theoretical expectations of increased teacher quality, regardless of previous empirical results.

Hanushek's (2003a) meta-analysis finds that over 40 percent of 'high quality' value-added model estimates find a positive and statistically significant relationship between teacher experience and student outcomes (p. F80). Some studies find connections between effectiveness and master's degree completion for math and science instructors whose graduate degree is in the substantive subject taught (as opposed to a

Masters in Education or a subject area unrelated to the course taught) (Goldhaber and Brewer, 1996). National Board Certification (Goldhaber and Anthony, 2005), higher educational institution quality (Ehrenberg and Brewer, 1994), and teacher test scores (Hanushek, 1997) also positively correlate with student achievement. The connection between highly qualified teacher status (defined in the NCLB[4] Act) may vary depending on the criteria used to achieve this status. The underlying teacher variables utilized in the study provide a more detailed measure of the specific attributes of high quality teachers compared to the less exact 'highly qualified teacher' status.

Measures of teacher characteristics exist at the classroom level. Teacher characteristics include dichotomous variables indicating teacher education level (master's or above), certification type other (includes emergency, provisional, and lateral licenses), National Board Certification status, in-field teacher, educational institution quality (based on Barron's rankings of undergraduate institutions that are 'Most Competitive' or 'Highly Competitive'), and a series of dichotomous variables representing various experience categories with five to nine years of experience functioning as the reference group. The model also includes continuous level variables based on test performance (a standardized average of all available teacher test scores) and total non-certified teacher pay. These payments include any ABCs program bonuses or local supplements. Certified pay is based on the years of experience, National Board Certification status, and the highest degree held by an instructor and does not vary across districts within the state on these factors.

[4] The No Child Left Behind Act of 2001 (NCLB) requires states to develop standards for identifying teachers as Highly Qualified Teachers and mandates that teachers in all core subject areas be 'highly qualified' by the end of the 2005-06 school year.

Classroom Characteristics

Within classroom controls include measures of the number of students within a classroom provided within the roster data, the percentage of students eligible for free or reduced lunch within a classroom, and the percentage of students within various racial/ethnic categories in each classroom. These measures are consistent with those utilized by Argys, et al. (1996) and Hanushek et al. (2003). In addition to those measures present in previous research, I include measures of curriculum differences between classrooms as indicated by course title keyword coding. This measure is one of the essential items necessary to properly estimate school tracking effects (Betts & Shkolnik, 2000b). Classrooms containing the words 'honors', 'advanced', or 'ap' are coded as advanced classrooms. Classrooms designated with names such as 'remedial', 'special', or 'resource' are coded as remedial curriculum classes (Honors Course Implementation Guide, 2005; Outline of the Course, 2008).

School Characteristics

Evidence for positive impacts of overall spending on student outcomes remains mixed. In meta-analytic studies conducted by Hanushek (1986; 1997), many instances of positive and negative relationships between spending levels and student outcomes are revealed depending on the resources considered and data available to researchers. Hanushek (1986; 1997) concludes in two studies that the preponderance of evidence supports no consistent relationship between school inputs measured in dollars of spending and student test score outcomes. Figlio (1999) and Sander (1999) find evidence of some positive impacts on student outcomes for schools which spend more on education provision utilizing alternative statistical methodologies. Figlio (1999) uses a translog production function as an alternative to standard EPF models and concludes that a positive relationship exists between school inputs and student outcomes measured as performance on NELS (National Educational Longitudinal Survey) and HBS (High School and Beyond) but does not identify the magnitude of these relationships. Sander (1999) finds a positive relationship between per pupil expenditures and eighth grade math scores in Illinois and identifies a meaningful effect size for this relationship.

Individual level compensation programs that provide bonuses or other incentives to teachers provide additional evidence on the connection between compensation and student outcomes. Clotfelter et al. (2006) find that a bonus program to retain math and science teachers in low-income or low-performing North Carolina schools had a positive impact on teacher turnover, but the link to positive student outcomes cannot be assessed due to discontinuation of the program. I include additional school-wide resources per student as a control for a school's overall expenditures in other classes. Finally, I use urban and rural school designations from the National Center for Education Statistics to create indicator variables representing schools in these two categories.

Additional school-level measures include variables which address the learning climate of the school. While work by Coleman, et al. (1966) addressed the issue of school climate as a function of racial composition, more recent studies have included school-level measures of income and school size to assess school climate. In this study, the percentage of students eligible for free or reduced lunch, percentage of students in various race/ethnicity categories within a school, rural/urban designation, teacher turnover percentage, and average daily membership (ADM) or school size function as control variables to adjust for differences in school enrollment and environment.

As a review, control variables included at the school level include interval level measures of expenditures per student depicting resource commitments to the overall school environment, school size, teacher turnover percentage, mean percentage of students in race/ethnicity categories, mean percentage of students eligible for free and reduced lunch, and dichotomous variables indicating rural or urban schools. As an alternative specification of tracked classrooms compared to previous research, I will compare the deviation in scores within classrooms to the deviation of scores expected with random assignment to determine the tracking intensity of schools. Table 3.3 provides a complete list of variables included in models.

Table 3.3 – List of Variables

Category	**Name**	**Description**	**Type**
Dependent Variable	EOC Standardized Outcome	Indicates standardized outcome score on EOC tests	Continuous
Variables of Interest	Peer Ability	Equal to the mean standardized value in current class based on 8th grade standardized scores (minus the influence of student i)	Continuous
	Low Performing Student	Scored in the bottom 20% in reading or math on 8th Grade EOG exam	Dichotomous
	High Performing Student	Scored in the top 20% in reading or math on 8th Grade EOG exam	Dichotomous
	Peer Ability * Low Performing Student	Variable interacting the Peer Ability variable with Low Performing Student status	Continuous
	Peer Ability * High Performing Student	Variable interacting the Peer Ability variable with High Performing Student status	Continuous
	Peer Dispersion (Classroom Level)	Continuous variable equal to the standard deviation of class prior achievement	Continuous
	Track Intensity (School Level)	Series of three indicator variables for differing track intensity	Dichotomous

Table 3.3 – List of Variables – Continued

Individual Level Measures			
	Male	Male indicator variable (reference group is Females)	Dichotomous
	Black Hispanic Other	Race/Ethnicity indicator variables (reference group is White)	Dichotomous
	Underage	Variable is 1 if student's birth date is before the cutoff date for the cohort	Dichotomous
	Overage	Variable is 1 if student's birth date is more than one year older than the cutoff date for the cohort (reference group is appropriate age for grade level)	Dichotomous
	Academically or Intellectually Gifted	Variable is 1 if student is coded as Gifted	Dichotomous
	Disabled Student	Variable is 1 if student is coded as exceptional (reference group is neither gifted nor disabled)	Dichotomous
	Free Lunch Reduced Lunch	Variable is 1 if student is coded as eligible for specified program (reference group is neither free or reduced lunch eligible)	Dichotomous
	PED Less than High School PED HS Graduate PED Bachelors or Higher	Variable is 1 if parental education is within the category (reference group is some college)	Dichotomous
	LEP Current Previous LEP Recipient	Variable is 1 if student is in specified Limited English Proficiency category (reference group never received LEP services)	Dichotomous
	Grade 10 Grade 11 Grade 12	Variable is 1 if student is in specified grade (reference group is grade 9)	Dichotomous
	Low-performing Student	Variable is 1 if student scored in bottom 20% of standardized scores in reading or math on 8th grade EOG test	Dichotomous

Table 3.3 – List of Variables – Continued

	High-performing Student	Variable is 1 if student scored in top 20% of standardized scores in reading or math on on 8th grade EOG test	Dichotomous
	Absences	Variable indicating the number of absences by the student during the school year	Continuous
Classroom Level Variables (Non-Teacher)	Students Per Classroom	Variable equal to the number of students within the class	Continuous
	Black Hispanic Other	Proportion of students within class in various race/ethnicity categories (reference group is white)	Continuous
	Free Lunch Reduced Lunch	Proportion of students within class eligible for free or reduced lunch (reference group is not free or reduced lunch eligible)	Continuous
	Advanced Curriculum Remedial Curriculum	Variable is 1 if course curriculum is in specified category (reference group is standard curriculum)	Dichotomous
Classroom Level Variables (Teacher)			
	Advanced Degree	Variable equal to 1 for teachers with Masters Degree or higher (reference group is bachelors teachers)	Dichotomous
	Licensed Other	Variable equal to 1 for teachers with temporary, emergency, provisional, or lateral licensure (reference group is initial or continuing license)	Dichotomous
	NBC Teachers	Variable equal to 1 for Nationally Board Certified teachers (reference group is non-NBC teachers)	Dichotomous

Table 3.3 – List of Variables – Continued

	Most or Highly Competitive Barron's Rating	Variable equal to 1 for teachers that graduated from a college rated as Most Competitive or Highly Competitive based on Barron's rating in 2004 (reference group is all other category ratings)	Dichotomous
	First Year Second Year Third Year Fourth or Fifth Year Sixth to Tenth Year Eleventh to Fifteenth Year Sixteenth to Twentieth Year Twenty First to Twenty Fifth Year Twenty-sixth Year Plus	Variable is 1 if teacher is in the specified experience category	Dichotomous
	Test Average	Mean standardized performance on any standardized test within the dataset (includes Praxis I or II, NTE, GRE, SAT, DPI tests)	Continuous
	Non-Certified Teacher Pay	Variable equal to the instructors pay from bonuses and local supplements in dollars	Continuous

Table 3.3 – List of Variables – Continued

School Level Variables	Total Per Pupil Expenditures (excluding Capital and Community Services)	Variable equal to the dollar amount of total per pupil expenditures in 100's of dollars	Continuous
	School Size (ADM)	Variable equal to the number of students enrolled in the school	Continuous
	Teacher Turnover	Variable equal to the percentage of teachers that did not return to the school from the prior (2003-04) school year	Continuous
	Black Hispanic Other	Proportion of students within school in various race/ethnicity categories (reference group is white)	Continuous
	Free Lunch Reduced Lunch	Proportion of students within school eligible for free or reduced lunch (reference group is not free or reduced lunch eligible)	Continuous
	Urban School Rural School	Value is 1 for schools in the designated category based on data from the National Center for Education Statistics (reference group includes suburban and town based schools)	Dichotomous

Modeling the Education Production Function

Equation 3.4 provides a simplified version of the fully specified model in ordinary least squares regression (OLS).

Equation 3.4

$$Y_{ics} = \beta_0 + \delta_{ics}Individual\&FamilyChar + \alpha_{cs}TeacherChar + \theta_{cs}ClassroomChar + \varphi_s SchoolChar + \varepsilon_{ics}$$

Explanatory variables within the production function appear across three ‘levels’ of measurement. These levels refer to the individual, classroom, or school level where variable measurement occurs. Table 3.4 provides a matrix, which places relevant independent variables into their appropriate level of measurement. The vector of individual characteristics, δ_{ics} includes the student, family, and peer achievement characteristics specified in the measurement section functioning at an individual level within the model. Within the same classroom, the characteristics of peer achievement vary across students and function within the vector of individual characteristics accordingly. The vectors of characteristics estimated by α_{cs} and θ_{cs} include teacher characteristics and classroom environment variables respectively, both measured at the classroom level. At the school level, φ_s includes the vector of school level measures of environment and context. Finally, ε_{ics} is an error term estimated for each individual normally distributed with a mean of zero and constant variance. Individual roster entries are weighted to total one observation for one individual’s EOC test score for English I and Biology models. For Algebra I models, weights are adjusted with the inverse mills ratio weighting. Specifics on the preparation of weights for Algebra I are found in Appendix A.

The three levels previously discussed (individual, classroom, and school) are nested levels that describe the organization of individuals grouped into classrooms and classrooms grouped within schools. While the term ε_{ics} describes the error associated with an individual level observation, the error has three components: 1) the variation among individuals within a classroom, 2) the variation between classrooms within schools, and 3) the variation across schools. Hierarchical Linear Model (HLM) is an extension of ordinary least squares (OLS) regression that possesses some distinct advantages over a traditional OLS model used to assess the influence of these variables on student outcomes (Raudenbush and Bryk, 2002). First, the HLM approach allows one to partition overall variance into components that represent the amount of influence that each level has on the predicted value of student outcomes.

Equation 3.5

$$Y_{ics} = \gamma_{000} + u_{00s} + r_{0cs} + e_{ics}$$

The intraclass correlation coefficient (ICC) describes the amount of variance in the dependent variable occurring within each of the three levels (individuals ($\mathrm{e_{ics}}$), classrooms ($\mathrm{r_{0cs}}$), or schools ($\mathrm{u_{00s}}$)). Equation 3.5 above presents a null model for the variance relationships among the three levels of the HLM model. In the null model, the dependent variable remains the same as that used in the overall model, the standardized test outcome for student *i* in classroom *c* in school *s*. When modeled, the null equation yields values for the four components. γ_{000} has the value of the mean dependent variable outcome, and the other three components are error terms representing variance at the three levels of the model: u_{00s} is variance between schools, r_{0cs} is variation between classrooms within schools, and e_{ics} is variation between individuals within classrooms. The square of the standard error for each of these three error terms is the variance component of each of the levels. To calculate the ICC, the variance of one of the terms is divided by the sum of the variances of all three terms (ie. ICC = $\frac{\mathrm{VAR}(e_{ics})}{(VAR(u_{00s}) + VAR(r_{0cs}) + VAR(e_{ics}))}$) (Kim 2009, p. 341). The ICC result example provides the proportion of total variance which occurs at the individual student level. Similarly, the between school ICC is given by $\frac{\mathrm{VAR}(u_{00s})}{(VAR(u_{00s}) + VAR(r_{0cs}) + VAR(e_{ics}))}$ (Hox 2002, Kim 2009). Using this approach and based on previous research (Bryk and Raudenbush, 1988), I expect that the majority of variation in student achievement occurs at the individual level with lesser amounts of variation occurring at the classroom and school levels.

Table 3.4 – Education Production Function Elements

		Student/Family	Teacher	Classroom	School	Variables of Interest
HLM Level	Level I – Individual	Prior Achievement (squared and cubed) Gender Race/Ethnicity Age Exceptionality Free/Reduced Lunch Status Parental Education Limited English Proficiency Grade Absences				Peer Ability Low Performing Student High Performing Student Interactions of Low and High Performing Student with Peer Ability
	Level II – Classroom		Advanced Degree Other License NBC Status Infield Teacher Barron's Quality Experience Test Performance Non-Certified Compensation	Students Per Classroom Proportion Free/Reduced Lunch Status Proportion Race/Ethnicity Adv/Remedial Curriculum		Peer Dispersion
	Level III – School				Per Pupil Exp. School Size Teacher Turnover Proportion Race/Ethnicity Proportion Free/Reduced Lunch Rural/Urban Schools	Tracking Intensity

HLM explicitly recognizes clusters of observations within data and corrects standard errors to account for the influence of clustering. These adjustments to standard errors actually increase the size of standard errors compared to OLS estimates and diminish the chances that specific tests of hypotheses are improperly deemed to be statistically significant beyond the level expected by chance alone.

Equation 3.6

$$Y_{ics} = \beta_0 + \beta_1 PeerAbility + \delta_{ics} IndividualChar + \alpha_{cs} TeacherChar + \theta_{cs} ClassroomChar + \varphi_s SchoolChar + u_{00s} + r_{0cs} + e_{ics}$$

Equation 3.6 is a reduced form equation representing the first model used in the analysis. Y_{ics} is the individual student's standardized outcome on an end of course exam, β_0 is the model intercept value, β_1 is the coefficient on the variable of interest (Peer Ability), δ_{ics} represents a vector of coefficients based on individual characteristic controls, α_{cs} represents a vector of coefficients based on teacher characteristics, θ_{cs} represents a vector of coefficients based on classroom characteristics, φ_s represents a vector of coefficients based on school characteristics for students within a school, and u_{00s}, r_{0cs}, and e_{ics} represent residual variance at the school, classroom, and individual level respectively.

Equation 3.7 expands on Equation 3.6 and displays the full mixed model used in the analysis of peer effects (Research Question 1). This equation provides the mixed model specification. The mixed model is a combination of separate equations at the three levels of analysis which are combined into a single equation via substitution. The model presented also includes the variable of peer ability, the variable of interest for Research

Question 1 below. Additional variables are added to models depending on the specific hypothesis under investigation.

Equation 3.7

$$
\begin{aligned}
Y_{ics} = \gamma_{000} &+ \gamma_{001} TotalPerPupilExp + \gamma_{002} ADM + \gamma_{003} Turnover \\
&+ \gamma_{004} SchoolPropByEthnicity + \gamma_{005} PropByFRLunch \\
&+ \gamma_{006} UrbanSchool + \gamma_{007} RuralSchool + \gamma_{010} StudentsInClass \\
&+ \gamma_{020} ClassPropByEthnicity + \gamma_{030} ClassPropByFRLunch \\
&+ + \gamma_{040} CurriculumAdvanced + \gamma_{050} CurriculumRemedial \\
&+ \gamma_{060} AdvDegree + \gamma_{070} LicensedOther + \gamma_{080} NBCTeacher \\
&+ \gamma_{090} InfieldTeacher + \gamma_{0100} HighBarron'sRating \\
&+ \gamma_{0110} TeacherExpCategories \\
&+ \gamma_{0120} TeacherTestAvg + \gamma_{0130} NonCertifiedTeacherPay \\
&+ \gamma_{100} PeerAbility + \gamma_{200} Std8thGradeMath \\
&+ \gamma_{300} Std8thGradeMathSq + \gamma_{400} Std8thGradeMathCubed \\
&+ \gamma_{500} Std8thGradeRead + \gamma_{600} Std8thGradeReadSq \\
&+ \gamma_{700} Std8thGradeReadCubed + \gamma_{800} Male + \gamma_{900} Ethnicity \\
&+ \gamma_{1000} UnderOrOverage + \gamma_{1100} AcademicallyGifted \\
&+ \gamma_{1200} Disabled + \gamma_{1300} FRLunch + \gamma_{1400} ParentalEducation \\
&+ \gamma_{1500} LEPStatus + \gamma_{1600} GradeLevel + \gamma_{1700} Absenses + u_{00s} + r_{0cs} \\
&+ e_{ics}
\end{aligned}
$$

Data

The dataset utilized for this study is a dataset organized for a study of the effectiveness of the Disadvantaged Student Supplemental Fund. Data files came directly from the North Carolina Department of Public Instruction and consisted of five primary sources:

1. Student-level roster data containing information on the subjects taken by each student and the instructor of record for the course for the 2004-2005 school year.
2. Student-level end of year test files for the 2000-01 through 2004-05 school years provide data on 8^{th} grade end of grade exams in the years 2000-01 to 2003-04 and end-of-course test results for high school students and student characteristics in 2004-05.
3. Teacher-level certification files provide data on experience, licensure, educational institution, and test performance.
4. Teacher-level compensation file provides data on teacher pay for the 2004-05 school year including local supplements and bonuses.
5. School and district-level expenditure file provides data on overall expenditure of resources organized by the state's Uniform Chart of Accounts.

The three primary types of data (student, teacher, and school expenditures) are linked through the most unique aspect of this data, the student roster information. Appendix B provides a more detailed discussion of the process used to link the roster to teacher specific data files. The linked data files contain information from all five sources arranged with a student roster entry and their current year's end-of-course test performance and personal characteristics, their previous 8th grade test performance, their specific teacher for the relevant course matched to their individual characteristics and pay information, and school level data on resources and school level characteristics. The number of roster entries for each student varies across schools and observations are weighted so that each student taking one individual EOC test is weighted as one. For example, a student who is enrolled in an Algebra I class for two semesters has each incidence of the course weighted one half in order to avoid having students with more listed course enrollments count multiple times.

Table 3.5 displays descriptive information on the number of cases used in models after compiling the complete data set for this study. Missing data are due to a number of factors including difficulty matching students to their 8th grades from up to four years before (students in grades 9 to 12), failures in matching between teacher names in roster and salary data during the roster matching phase, missing values in teacher characteristics which include teacher test score information, and a low rate of missing values for school characteristics. Missing classroom characteristics increase the missing data rate substantially since a single missing value for a classroom eliminates all student observations within that classroom from the analysis. Including the days absent variable for student attendance also increased the rate of missing values at the student level. Overall, about 55 percent of students testing in English I and Biology are included in the models along with over 66 percent of students for whom inverse mills ratio weights could be calculated in Algebra I.

Table 3.5 – Cases Lost Due to Missing Data

	English I	Algebra I	Biology
Initial Observations in Dataset	125,948	100,554	97,327
Initial Individual Students Represented	102,803	47,032	77,869
% Missing Grade 8 Scores	11.7%	0.0%	12.1%
% Missing Any Student Characteristic	28.5%	11.0%	29.3%
% Missing Classroom Characteristic	20.7%	22.7%	21.3%
% Missing School Characteristic	5.0%	4.5%	3.2%
% Listwise Missing	44.3%	33.3%	44.6%
Final Observations in Dataset	69,158	68,352	53,098
Final Individual Students Represented	57,257	31,393	43,101

Biology and English I provide the widest cross section of students available in the high school dataset for use in this study. In North Carolina, students not obtaining proficient scores on Algebra I and English I on end-of-course (EOC) exams must sit for the state's High School Comprehensive Test[5]. This policy appears to influence enrollment in Algebra I and English I compared to Biology classrooms. The next section describes the data utilized in models and highlights the differences in student, classroom, and school characteristics across the three subjects.

[5] http://www.ncpublicschools.org/docs/accountability/NORTHCgeneralpolicies.pdf accessed on July 2, 2007.

Descriptive Statistics

Appendix C provides descriptive statistics from the three areas of the study – student achievement in English I, Algebra I and Biology. Descriptive statistics are organized with individual level data first followed by classroom and school level means, standard deviations, minimums, and maximums. The largest sample of students included in the analysis is in English I. Nearly 70,000 case observations represent about 57,200 students taking English I in the 2004-05 school year. The Biology analysis represents about 43,100 students, and Algebra I contains 31,393 individual student observations.

The prior ability level of peers is lower in Algebra I than the other two subjects, as expected based on the selection into 8th grade Algebra I by higher performing students. Biology students have substantially higher 8th grade test scores, nearly 1 standard deviation above the mean in both reading and math, than English I students, scoring on average about 0.1 standard deviations above the mean in both reading and math, followed by Algebra I students whose average grade 8 reading and math scores are about 0.2 standard deviations *below* the mean. On average, students taking Algebra I in high school grades are more likely to be male, black, overage for their grade level, free lunch eligible, and to be classified as low performing students compared to students enrolled in English I and Biology. They are also less like to be white or classified as high performing students. Students enrolled in Algebra I and English I are primarily enrolled in 9th grade while Biology students are primarily in 10th grade.

Fewer classrooms are represented in the Biology analysis than are included for Algebra I and English I. This difference is attributed primarily due to differences in the roster listings of Biology which report fewer entries per student per test and slightly larger class sizes in Biology (about 20.7 students per class) versus less than 20 students per class in both Algebra I and English I. Comparing the classroom characteristics in each of the three subjects finds more variation than was evident in school characteristics. Classroom peer ability dispersion is higher in English I and Biology compared to Algebra I classrooms. Given the selection of higher performing students into Algebra I in 8th grade, it is no surprise to find that the mean peer ability dispersion value for Algebra I, 0.61, is lower than the values for both English I, 0.69, and Biology, 0.73. Biology classrooms contain slightly higher percentages of white students and lower percentages of free lunch eligible students compared to English I and Algebra I classrooms. The importance of a wide variety of control variables for this type of analysis is emphasized by the variance across subjects in curriculum level offerings in the three subjects. Less than ten percent of Algebra I classrooms are coded as advanced curriculum compared to more than a quarter of classrooms in Biology and English I. Also, six percent more classrooms in Algebra I are remedial compared to Biology and English I.

Across the three subjects, teacher characteristics for high school Algebra I students are taught by teachers that are on average less likely to be nationally board certified, less likely to be teaching with an initial or continuing license in the subject being taught (infield teachers), and have lower scores on tests of general academic ability or subject knowledge than teachers teaching English I or Biology. Other teacher characteristics are fairly similar across subjects.

Comparing the three subjects, school characteristics are very similar across the three subjects with the exclusion of the total number of schools. The Biology data set represents 258 unique schools while the English I and Algebra I data sets include 287 and

286 schools respectively. Having reviewed the descriptive characteristics of the three data sets, the next section presents the research questions pursued in this analysis.

Research Questions

This study focuses on testing three sets of questions representing the relationships between (1) peer ability, (2) peer dispersion, and (3) school level ability tracking and student end-of-course test score outcomes. The outcome variable utilized is consistent across hypotheses and is the student's standardized end-of-course (EOC) test score on exams in English I, Algebra I, or Biology.

Research Question 1 (RQ1): Do student test scores increase more in classrooms where peers, on average, have higher prior achievement scores or averages? (Peer Ability)

(Peer ability has a positive and significant relationship to test score gains.)

Zimmer and Toma (2000) provide empirical support for this hypothesis in their work on peer influences across countries. Theoretical support for an efficiency argument and peer co-operative learning at a high level due to increased peer skills should work in concert to provide increased test score gains for students in classrooms with higher-achieving peers.

Equation 3.6 (pg. 70) tests RQ1 through the inclusion of the $PeerAbility$ variable, the mean performance of peers on standardized end of grade tests in math or reading during grade 8, depending on the subject under study in the model. If RQ1 is true, the coefficient, β_1, will be positive and statistically significant indicating that students with higher achieving peers score higher on their own end of course tests in specific subjects.

Equation 3.8

$$Y_{ics} = \beta_0 + \beta_1 PeerAbility + \beta_2 PeerAbilitySquared + \beta_3 PeerAbilityCubed + \delta_{ics} IndividualChar + \alpha_{cs} TeacherChar + \theta_{cs} ClassroomChar + \varphi_s SchoolChar + u_{00s} + r_{0cs} + e_{ics}$$

Equation 3.8 adds terms to the original peer effects model (Equation 3.6) to test for non-linear relationships between peer ability and a student's observed EOC test score outcome. These terms allow for a diminishing effect in any observed relationship between peer ability and individual student test score outcomes. Statistically significant values of β_2 and β_3 would indicate a non-linear relationship between peer academic ability and student test score outcomes.

> Research Question 2 (RQ2): Do student test scores increase more in classrooms with more heterogeneous ability grouping? (Peer Dispersion)
>
> (Heterogeneity measured as the standard deviation of classroom peers has a positive, significant relationship to test score gains.)

Research Question 2 directly assesses the impact of peer ability dispersion on student test scores. Three studies find a positive relationship between peer ability dispersion and student test score outcomes (Vigdor and Nechyba, 2004; Zabel, 2008; Zimmer and Toma, 2000), while another finds no statistically significant relationship between these variables (Hanushek, et al., 2003). This hypothesis will determine the relationship for students in high school English I, Algebra I, and Biology.

Equation 3.9

$$Y_{ics} = \beta_0 + \delta_{ics} IndividualChar + \beta_1 PeerDispersion + \alpha_{cs} TeacherChar + \theta_{cs} ClassroomChar + \varphi_s SchoolChar + u_{00s} + r_{0cs} + e_{ics}$$

If classrooms with higher ability diversity are correlated with increased student performance, term β_1 will be positive and statistically significant. If significant, the model will be extended with a squared term based on the peer dispersion. A positive and significant coefficient on this variable has important policy implications for student equity. While few critics have argued that tracking is inefficient, a positive relationship between increased student diversity and student test score outcomes would mean decreased diversity has negative consequences for student achievement.

Equation 3.10

$$Y_{ics} = \beta_0 + \beta_1 PeerAbility + \beta_2 LowPerformingStudent + \beta_3 HighPerformingStudent + \beta 4 PeerAbility * LowPerformingStudent + \beta 5 PeerAbility * HighPerformingStudent + \delta_{ics} IndividualChar + \alpha_{cs} TeacherChar + \theta_{cs} ClassroomChar + \varphi_s SchoolChar + u_{00s} + r_{0cs} + e_{ics}$$

In addition to models testing the relationship between peer ability and peer dispersion on student outcomes, I test whether peer effect relationships differ for students in the top or bottom 20% of the test score distribution based on 8th grade EOG tests. This follows the high and low performing student definitions utilized by Zimmer and Toma (2000). Equation 3.10 provides an example of a model utilizing indicator variables for high and low performing students and interactions between these two variables and the individual's peer ability measure. Differential outcomes for high or low performing students are based on statistical significance tests for the terms β_4 andβ_5, where statistical significance indicates a differential outcome for high or low performing students respectively.

Research Question 3 (RQ3): Do tracked classroom configurations generate significantly different outcomes compared to heterogeneous classroom configurations? (Tracking)

(Total aggregate student gains differ when student ability grouping is utilized within a school.)

Research Question 3 directly assesses the differences in outcomes due to school-wide ability tracking by testing the relationship between a school's ability tracking intensity and student EOC test score outcomes.

Equation 3.9

$$\begin{aligned} Y_{ics} = \beta_0 + \beta_1 HighTrackedSchool + \beta_2 ModerateTrackedSchool \\ + \beta_3 LowTrackedSchool + \delta_{ics}^{*} IndividualChar^{*} + \alpha_{cs}^{*} TeacherChar^{*} \\ + \varphi_s SchoolChar + u_{00s} + r_{0cs} + e_{ics} \end{aligned}$$

In addition to the creation of tracking definition variables, I reduce the number of variables at the classroom and individual level by removing those characteristics that might plausibly explain a relationship between tracking and achievement. The * designates that a limited number of control variables are included in these total effects models. For example, peer effects might mediate differences in outcomes in tracked versus un-tracked schools. Also, principals may place more able teachers with either very high or very low ability students in order to generate higher aggregate test score gains for schools of a given tracking intensity. I remove peer ability measures at the individual level and peer dispersion, curriculum level indicators, and teacher characteristics from the classroom level for the initial models testing tracking differences. I add these variables to subsequent models if the results show associations between the tracking variables and outcomes to test mediating relationships between these variables.

Equation 3.11

$$Y_{ics} = \beta_0 + \beta_1 HighIntenstiyTracking + \beta_2 ModerateIntenstiyTracking \\ + \beta_3 LowIntenstiyTracking + \delta_{ics} IndividualChar \\ + \alpha_{cs} TeacherChar + \theta_{cs} ClassroomChar + \varphi_s SchoolChar + u_{00s} \\ + r_{0cs} + e_{ics}$$

Equation 3.11 presents a reduced form model of the equation testing the relationship between tracking intensity and EOC test score outcomes which includes possible mediating variables that might explain any differences between tracked and un-tracked classrooms. For example, if the inclusion of the peer ability variables reduces the tracking indicator variables to statistical insignificance and the peer ability variables are statistically significant, this would indicate that the relationship between tracking and student outcomes is solely a function of changes to classroom peer groupings. Moves toward zero of β_1, β_2, or β_3, in combination with statistically significant mediator variables indicate the mechanisms by which tracking intensity changes predict student outcomes. Three types of mediators are tested when any tracking variables are statistically significant: peer ability, curriculum differences, and teacher characteristics. I add each group of variables in single models and one combined model if multiple mediators appear to impact the observed relationship between tracking intensity and outcomes. Chapter 4 provides a discussion of modeling results.

CHAPTER 4

MODEL RESULTS

Introduction

This chapter presents the results of models designed to estimate the impact of peer effects and student tracking on student end-of-course test score outcomes in three high school subjects. The purpose of the analysis is to understand how variation in the ability levels of peers in one's classroom, based on prior test score performance, influences test score outcomes. In addition, this analysis seeks to understand how tracking policies designed to restrict prior test performance variability affect student achievement. These three courses, English I, Algebra I, and Biology, have broad enrollment for high school students in North Carolina due to their inclusion as requirements for the three primary plans of study in North Carolina's public schools. This study has the potential to reveal how the relationship between student grouping and test score outcomes varies across these subjects. End-of-course exams are required for all students enrolled in any of these three subjects and these courses are required for graduation from North Carolina High Schools in the *Career Prep, College Tech Prep,* and *College/University Prep* plans of study (High School Graduation Requirements). Some students with cognitive disabilities are permitted to enroll in the *Occupational* course of study which does not include EOC testing required courses.

The results are first presented by subject and are organized around the three primary research questions presented in Chapter 3. The analyses specifically focus on the following hypotheses:

1. RQ1: Student test scores increase more in classrooms where peers, on average, have higher prior achievement scores or averages. (Peer Ability)
2. RQ_2: Student test scores increase more in classrooms with more heterogeneous ability grouping. (Peer Dispersion)
3. RQ_3: Tracked classroom configurations generate significantly different outcomes compared to heterogeneous classroom configurations. (Tracking)

While the focus of the study is directly addressing these three hypotheses, I will present additional models investigating non-linear relationships and differential effects for high and low performing students based on a student's prior test score. For the tracking analysis, I will also present additional models focused on identifying variables that mediate any observed relationship between tracked classrooms and student test score outcomes.

The chapter is organized around the three subjects analyzed for this study, English I, Algebra I, and Biology. For each subject, I present the series of models used to evaluate the previously presented hypotheses along with a presentation of the key control variables included in models of peer effects. After examining each of the subjects in turn, I conclude with a presentation of the results across subjects.

English I Findings

An initial null model which allows the isolation of variance across the three modeling levels reveals that in English I, variance is largely restricted to the student and classroom level (Table 4.1). School level grouping is only responsible for about six percent of the total variation in student test scores in a null model where only the groupings of students into classrooms and schools are considered as predictors in the model. As expected, the intraclass correlation coefficients indicate that most variation occurs at the student level with significant additional variation at the classroom level.

Table 4.1 - English I – ICC Calculation

	Variance	Intraclass Correlation Coefficient
Level I - Student	0.497	49.7%
Level II - Classroom	0.441	44.1%
Level III - School	0.063	6.3%

Table 4.2 - English I Results

		(1)		(2)		(3)	
		Base Model - E1 No Peer Effects		Model - E1 With Peer Effects - Reading		Model - E1 With Peer Effects - Reading Squared and Cubed	
Coefficient Group		Coef.	Std. Err	Coef.	Std. Err	Coef.	Std. Err
School-Level	Intercept	0.052	(0.020)*	0.089	(0.02)**	0.086	(0.019)**
	Tracking Quartile - Lowest	----	----	----	----	----	----
	Tracking Quartile - Medium	----	----	----	----	----	----
	Tracking Quartile - Highest	----	----	----	----	----	----
	Total Per Pupil Exp (excl. Capital & Com. Ser.)	0.002	(0.000)	0.002	(0.000)	0.002	(0.000)
	Average Daily Membership	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
	Teacher Turnover	0.001	(0.000)	0.002	(0.000)	0.002	(0.000)
	School Pct Black	0.001	(0.000)	0.002	(0.000)**	0.002	(0.000)**
	School Pct Hispanic	0.004	(0.002)*	0.005	(0.001)*	0.005	(0.001)*
	School Pct Other	0.001	(0.001)	0.000	(0.001)	0.000	(0.001)
	School Pct Free Lunch	-0.004	(0.001)**	-0.003	(0.000)**	-0.003	(0.000)**
	School Pct Reduced Lunch	-0.004	(0.003)	-0.004	(0.003)	-0.004	(0.003)
	Urban Area School	-0.017	(0.021)	-0.018	(0.020)	-0.020	(0.020)
	Rural Area School	0.004	(0.015)	0.006	(0.014)	0.007	(0.014)

Table 4.2 - English I Results - Continued

Classroom-Level							
	Peer Dispersion - Reading (Std. Dev.)	----	----	----	----	----	----
	Number of Students in Class	0.005	(0.000)**	0.003	(0.000)**	0.003	(0.000)**
	Classroom Percent Black	-0.001	(0.000)	0.000	(0.000)	0.000	(0.000)
	Classroom Percent Hispanic	-0.002	(0.000)**	0.000	(0.000)	0.000	(0.000)
	Classroom Percent Other	0.001	(0.000)	0.001	(0.000)	0.001	(0.000)
	Classroom Percent Free Lunch Eligible	0.000	(0.000)	0.001	(0.000)	0.000	(0.000)
	Classroom Percent Reduced Lunch Eligible	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
	Curriculum Advanced	0.162	(0.012)**	0.094	(0.013)**	0.082	(0.015)**
	Curriculum Remedial	-0.002	(0.013)	0.002	(0.012)	0.001	(0.012)
	Advanced Degree	-0.002	(0.009)	-0.003	(0.008)	-0.004	(0.008)
	Licensed Other	0.043	(0.017)*	0.029	(0.017)	0.030	(0.017)
	National Board Certification	0.013	(0.013)	0.015	(0.013)	0.014	(0.013)
	Infield Teacher	0.033	(0.012)*	0.015	(0.012)	0.017	(0.012)
	Most or Highly Competitive Barron's Rating	-0.001	(0.011)	-0.003	(0.010)	-0.004	(0.010)
	First Year Teacher	-0.032	(0.016)*	-0.029	(0.015)	-0.030	(0.015)
	1 to 2 Years Teacher Experience	0.001	(0.014)	0.002	(0.013)	0.001	(0.013)
	3 to 4 Years Teacher Experience	-0.014	(0.014)	-0.015	(0.013)	-0.016	(0.013)
	10 to 14 Years Teacher Experience	0.014	(0.014)	0.008	(0.014)	0.006	(0.014)
	15 to 19 Years Teacher Experience	0.021	(0.017)	0.013	(0.016)	0.013	(0.016)
	20 to 24 Years Teacher Experience	0.012	(0.019)	0.007	(0.018)	0.007	(0.018)

Table 4.2 - English I Results - Continued

	25 or more Years Teacher Experience	0.033	(0.014)*	0.022	(0.014)	0.020	(0.014)
	Teacher Test Avg (Praxis, NTE, etc.)	-0.003	(0.007)	-0.005	(0.007)	-0.006	(0.007)
	Teacher Non-Certified Pay	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
Individual-Level							
	Peer Ability Math	----	----	----	----	----	----
	Peer Ability Reading	----	----	0.122	(0.012)**	0.124	(0.015)**
	Peer Ability Reading Squared	----	----	----	----	0.016	(0.006)*
	Peer Ability Reading Cubed	----	----	----	----	0.004	(0.005)
	Low-Performing Student * Peer Ability Reading	----	----	----	----	----	----
	High-Performing Student * Peer Ability Reading	----	----	----	----	----	----
	Std Math Score (Grade 8)	0.211	(0.005)**	0.207	(0.005)**	0.207	(0.005)**
	Std Math Score Squared (Grade 8)	0.006	(0.002)*	0.006	(0.002)*	0.006	(0.002)*
	Std Math Score Cubed (Grade 8)	0.000	(0.001)	0.000	(0.001)	0.000	(0.001)
	Std Read Score (Grade 8)	0.527	(0.006)**	0.525	(0.006)**	0.525	(0.006)**
	Std Read Score Squared (Grade 8)	-0.014	(0.002)**	-0.013	(0.002)**	-0.013	(0.002)**
	Std Read Score Cubed (Grade 8)	-0.022	(0.001)**	-0.022	(0.001)**	-0.022	(0.001)**
	Male	-0.149	(0.004)**	-0.148	(0.004)**	-0.147	(0.004)**
	Black	-0.080	(0.007)**	-0.083	(0.007)**	-0.084	(0.007)**
	Hispanic	-0.001	(0.014)	-0.006	(0.014)	-0.006	(0.014)
	Other	-0.023	(0.011)	-0.024	(0.011)*	-0.024	(0.011)*

Table 4.2 - English I Results - Continued

	Underage Student based on Grade	0.070	(0.018)**	0.069	(0.018)**	0.069	(0.018)**
	Overage Student based on Grade	-0.087	(0.007)**	-0.086	(0.007)**	-0.087	(0.007)**
	Academically or Intellectually Gifted	0.126	(0.008)**	0.121	(0.008)**	0.120	(0.008)**
	Disabled Student	-0.261	(0.011)**	-0.243	(0.011)**	-0.245	(0.011)**
	Free Lunch	-0.035	(0.006)**	-0.037	(0.006)**	-0.037	(0.006)**
	Reduced Lunch	-0.007	(0.008)	-0.008	(0.008)	-0.008	(0.008)
	Parent Education Less than High School	-0.028	(0.010)**	-0.027	(0.010)**	-0.027	(0.010)**
	Parent Education Some College	0.062	(0.005)**	0.061	(0.005)**	0.061	(0.005)**
	Parent Education College Graduate	0.074	(0.006)**	0.072	(0.006)**	0.072	(0.006)**
	Parent Education Missing	-0.107	(0.032)**	-0.112	(0.032)**	-0.112	(0.032)**
	Previous LEP Services Recipient	-0.054	(0.024)*	-0.053	(0.024)*	-0.053	(0.024)*
	LEP Services Recipient	-0.145	(0.020)**	-0.140	(0.020)**	-0.140	(0.020)**
	Grade 10	0.061	(0.030)*	0.065	(0.030)*	0.064	(0.030)*
	Grade 11	0.232	(0.090)*	0.232	(0.089)*	0.232	(0.090)*
	Grade 12	0.320	(0.152)*	0.312	(0.153)*	0.312	(0.153)*
	Days Absent	-0.005	(0.000)**	-0.005	(0.000)**	-0.005	(0.000)**
	Low-Performing Student	----	----	----	----	----	----
	High-Performing Student	----	----	----	----	----	----

Table 4.2 - English I Results – Continued

		(4)		(5)		(6)	
		Model - E1 With Peer Effects - Reading Only - Peer Dispersion		Model - E1 With Peer Effects - Reading Only - Peer Dispersion - Low and High Performing Student Interactions		Model - E1 With Peer Effects - Reading Only - Tracking	
Coefficient Group		Coef.	Std. Err	Coef.	Std. Err	Coef.	Std. Err
School-Level	Intercept	0.115	(0.025)**	0.089	(0.019)**	0.129	(0.018)**
	Tracking Quartile - Lowest	----	----	----	----	0.005	(0.017)
	Tracking Quartile - Medium	----	----	----	----	-0.023	(0.016)
	Tracking Quartile - Highest	----	----	----	----	-0.031	(0.017)
	Total Per Pupil Exp (excl. Capital & Com. Ser.)	0.001	(0.000)	0.002	(0.000)	0.002	(0.000)*
	Average Daily Membership	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)*
	Teacher Turnover	0.002	(0.000)	0.002	(0.000)	0.002	(0.000)
	School Pct Black	0.002	(0.000)**	0.002	(0.000)**	0.001	(0.000)
	School Pct Hispanic	0.005	(0.001)*	0.005	(0.001)*	0.003	(0.001)
	School Pct Other	0.000	(0.001)	0.000	(0.001)	0.001	(0.001)
	School Pct Free Lunch	-0.003	(0.000)**	-0.003	(0.000)**	-0.004	(0.001)**
	School Pct Reduced Lunch	-0.004	(0.003)	-0.004	(0.003)	-0.007	(0.003)*
	Urban Area School	-0.019	(0.020)	-0.020	(0.020)	-0.001	(0.022)
	Rural Area School	0.007	(0.014)	0.007	(0.014)	0.006	(0.015)

Table 4.2 - English I Results – Continued

Classroom-Level							
	Peer Dispersion - Reading (Std. Dev.)	-0.041	(0.023)	----	----	----	----
	Number of Students in Class	0.003	(0.000)**	0.003	(0.000)**	0.006	(0.000)**
	Classroom Percent Black	0.000	(0.000)	0.000	(0.000)	-0.002	(0.000)**
	Classroom Percent Hispanic	0.000	(0.000)	0.000	(0.000)	-0.003	(0.000)**
	Classroom Percent Other	0.001	(0.000)	0.001	(0.000)	0.000	(0.000)
	Classroom Percent Free Lunch Eligible	0.001	(0.000)	0.001	(0.000)	-0.002	(0.000)**
	Classroom Percent Reduced Lunch Eligible	0.000	(0.000)	0.000	(0.000)	-0.001	(0.000)**
	Curriculum Advanced	0.081	(0.015)**	0.081	(0.015)**	----	----
	Curriculum Remedial	0.001	(0.012)	0.002	(0.012)	----	----
	Advanced Degree	-0.004	(0.008)	-0.004	(0.008)	----	----
	Licensed Other	0.030	(0.017)	0.030	(0.017)	----	----
	National Board Certification	0.015	(0.013)	0.015	(0.013)	----	----
	Infield Teacher	0.017	(0.012)	0.018	(0.012)	----	----
	Most or Highly Competitive Barron's Rating	-0.003	(0.010)	-0.003	(0.010)	----	----
	First Year Teacher	-0.029	(0.015)	-0.029	(0.015)	----	----
	1 to 2 Years Teacher Experience	0.001	(0.013)	0.001	(0.013)	----	----
	3 to 4 Years Teacher Experience	-0.015	(0.013)	-0.016	(0.013)	----	----
	10 to 14 Years Teacher Experience	0.007	(0.014)	0.007	(0.014)	----	----
	15 to 19 Years Teacher Experience	0.013	(0.016)	0.013	(0.016)	----	----
	20 to 24 Years Teacher Experience	0.007	(0.018)	0.007	(0.018)	----	----

Table 4.2 - English I Results – Continued

	25 or more Years Teacher Experience	0.021	(0.014)	0.021	(0.014)	----	----
	Teacher Test Avg (Praxis, NTE, etc.)	-0.006	(0.007)	-0.006	(0.007)	----	----
	Teacher Non-Certified Pay	0.000	(0.000)	0.000	(0.000)		
Individual-Level							
	Peer Ability Math	----	----	----	----	----	----
	Peer Ability Reading	0.126	(0.012)**	0.134	(0.013)**	----	----
	Peer Ability Reading Squared	0.011	(0.006)	0.014	(0.008)	----	----
	Peer Ability Reading Cubed	----	----	----	----	----	----
	Low-Performing Student * Peer Ability Reading	----	----	-0.002	(0.018)	----	----
	High-Performing Student * Peer Ability Reading	----	----	-0.010	(0.013)	----	----
	Std Math Score (Grade 8)	0.206	(0.005)**	0.208	(0.006)**	0.218	(0.005)**
	Std Math Score Squared (Grade 8)	0.006	(0.002)*	0.008	(0.003)**	0.008	(0.002)**
	Std Math Score Cubed (Grade 8)	0.000	(0.001)	0.000	(0.001)	-0.001	(0.001)
	Std Read Score (Grade 8)	0.526	(0.006)**	0.525	(0.007)**	0.535	(0.006)**
	Std Read Score Squared (Grade 8)	-0.013	(0.002)**	-0.012	(0.002)**	-0.013	(0.002)**
	Std Read Score Cubed (Grade 8)	-0.022	(0.001)**	-0.022	(0.001)**	-0.023	(0.001)**
	Male	-0.147	(0.004)**	-0.147	(0.004)**	-0.154	(0.004)**
	Black	-0.083	(0.007)**	-0.084	(0.007)**	-0.076	(0.007)**
	Hispanic	-0.006	(0.014)	-0.006	(0.014)	0.002	(0.014)
	Other	-0.024	(0.011)*	-0.024	(0.011)*	-0.021	(0.011)
	Underage Student based on Grade	0.069	(0.018)**	0.069	(0.018)**	0.072	(0.019)**

Table 4.2 - English I Results – Continued

	Overage Student based on Grade	-0.086	(0.007)**	-0.086	(0.007)**	-0.089	(0.007)**
	Academically or Intellectually Gifted	0.120	(0.008)**	0.122	(0.008)**	0.142	(0.007)**
	Disabled Student	-0.244	(0.011)**	-0.245	(0.011)**	-0.263	(0.011)**
	Free Lunch	-0.037	(0.006)**	-0.037	(0.006)**	-0.031	(0.006)**
	Reduced Lunch	-0.008	(0.008)	-0.008	(0.008)	-0.005	(0.008)
	Parent Education Less than High School	-0.026	(0.010)**	-0.026	(0.010)*	-0.029	(0.010)**
	Parent Education Some College	0.061	(0.005)**	0.061	(0.005)**	0.064	(0.005)**
	Parent Education College Graduate	0.072	(0.006)**	0.072	(0.006)**	0.079	(0.006)**
	Parent Education Missing	-0.113	(0.032)**	-0.113	(0.032)**	-0.103	(0.032)**
	Previous LEP Services Recipient	-0.053	(0.024)*	-0.053	(0.024)*	-0.054	(0.024)*
	LEP Services Recipient	-0.141	(0.020)**	-0.140	(0.020)**	-0.139	(0.020)**
	Grade 10	0.064	(0.030)*	0.064	(0.030)*	0.059	(0.030)
	Grade 11	0.231	(0.089)*	0.232	(0.090)*	0.228	(0.091)*
	Grade 12	0.313	(0.152)*	0.313	(0.153)*	0.320	(0.150)*
	Days Absent	-0.005	(0.000)**	-0.005	(0.000)**	-0.005	(0.000)**
	Low-Performing Student	----	----	-0.011	(0.010)	----	----
	High-Performing Student	----	----	-0.011	(0.008)	----	----

Peer Ability and English I Achievement

Research Question 1 predicts that students with higher performing peers (measured by averaging the 8th grade reading test performance of all of the other students in a classroom) will score better on end-of-course English I exams compared to similar students in classrooms with lower performing peers. RQ 1 is supported by the positive and statistically significant coefficient on Peer Ability Reading in English I with the model results ‘E1 with Peer Effects – Reading’ (Table 4.2 Model (2)). Placing a student in a class where one’s peers scored one standard deviation higher than average is expected to increase that student’s performance on the EOC English I exam by about 0.122 standard deviations, controlling for other student, classroom, and school characteristics. Students placed in classrooms with lower performing peers are expected to perform more poorly by the same margin.

The relationship between the prior reading test performance of peers and current end-of-course English I test performance may be non-linear, and an additional model adds squared and cubed terms on peer ability to check for the proper functional form of the relationship between these two variables. For example, as prior peer performance increases, the expected increase in current test score performance may increase by a ratio that is not one-to-one. Model 3 (Table 4.2 Model (3)) tests for non-linear impacts of peer reading ability on a student’s English I test performance. A non-linear relationship between prior peer reading test performance and English I end-of-course test performance is supported by the positive and statistically significant Peer Ability Reading variable in model (3) (Table 4.2). Based on this model, the influence of peer ability on English I test performance increases at an increasing rate over the range of relevant peer ability scores. For a student with peers whose scores were 2 standard deviations above average, the predicted value of the student's English I score is an increase of about 0.31 standard deviations compared to an increase of only 0.14 standard deviation units for a similar student whose peers scored only one standard deviation above average. To summarize, the relationship between peer reading ability and end-of-course English I test scores appears to be positive and increasing at an increasing rate supporting Research Question 1 in English.

<u>Peer Dispersion and English I Achievement</u>

Just as higher performing peers may increase the learning gains of students, having too narrow a range of abilities within a classroom may limit the ability of students to learn from other students within a classroom. Research Question 2 predicts that students in more academically diverse classrooms (those where the dispersion of prior reading test performance is greater) will score better on English I exams compared to similar students. Model (4) provides a test of this research question by including a measure of how much student prior test performance varies within classrooms (Peer Dispersion – Reading (Std. Dev.)) (Table 4.2). The results of Model (4) suggest that RQ 2 is rejected as there is no statistically significant difference in test performance between similar students in classrooms that are more or less academically dispersed based on prior reading test performance.

Finally, we assess the extent to which the previously observed non-linear relationship between prior peer reading performance and English I test performance varies for students in the top and bottom quintiles based on 8^{th} grade test score performance in reading or math. The marginal returns to peer ability may differ for particularly high or low performing students based on their test score performance in grade eight. Model (5) adds indicator variables for students who scored in the top or bottom 20 percent of the distribution of test takers based on their end-of-grade test performance in 8th grade reading or math and also interacts these variables with the Peer Ability Reading value to look for unique relationships between academically disadvantaged students and peer characteristics (Table 4.2).[6] In model (5), I find no statistically significant difference in the influence of peer ability on student outcomes for high or low performing students in English I. The returns to enhanced peer ability appear consistent regardless of where the student scored on 8th grade end-of-grade exams in reading and math. For English I students, I reject the hypothesis that more heterogeneous ability groupings are associated with higher test score performance and further find no unique relationship between academically disadvantaged students and English I test score performance and peer ability.

Tracking and English I Achievement

While the prior two research questions deal with the relationship between peer and classroom characteristics and individual student test score performance, Research Question 3 focuses on how schools might choose to arrange students within classrooms based on their prior test score performance. Schools were arranged into four quintiles based on the differences in observed classroom characteristics, and the classroom characteristics predicted through a random assignment process of students to classrooms.

[6] Interactions with non-linear Peer Ability variables are excluded during these model runs to limit the possibility that statistically insignificant values are the result of multi-collinearity and not a lack of relationship between independent and dependent variables.

As classrooms become more tracked (or sorted) based on prior 8th grade test performance, the difference between the observed standard deviation of within classroom prior reading test performance and the value of randomly assigned students to classrooms standard deviation of within classroom prior reading test performance becomes more negative. High intensity tracked classrooms on the basis of prior test score performance have narrow classroom level standard deviation values as students within a classroom have very similar prior test scores. Random assignment of students to classrooms generates higher prior test score performance standard deviations. The lowest quartile of differences mimicked random assignment and this group of schools (non-tracked or un-tracked schools) serves as the reference group for the analysis. The three upper quartiles are arranged based on how intensely students are 'tracked' based on prior 8th grade reading test performance. The lowest quartile is most similar to non-tracked schools and the highest quartile schools are those in which there is the greatest difference between random assignment and observed classroom assignments based on the student's 8th grade end-of-grade reading test performance.

Research Question 3 investigates the link between tracking intensity at the high school level and student outcomes on end-of-course exams in English I. These models exclude regression coefficients related to the ability level of peers, curriculum level of the class (advanced, remedial, or regular), and teacher characteristics as these three groups of coefficients serve as possible mediators of the impact of tracking intensity on student achievement. The three mediator models are implemented if a 'gross' effect of student tracking is revealed in the initial tracking model. Model (6) includes three dichotomous tracking intensity indicator variables at the school level with un-tracked schools serving as the reference group (Table 4.2). For English I students, no relationship is observed between the intensity of tracking within a school and student test score outcomes when controlling for student, classroom, and school characteristics not expected to be correlated with tracking. For English I, no additional tracking models are implemented as I reject the hypothesis that there is a relationship between tracking intensity based on student reading test performance and English I end-of-course test performance (Table 4.2 Column 6).

English I Control Variables

The base model (Model 1) for English I students shows the expected strong influence of individual characteristics on student outcomes and lesser impacts of characteristics measured at the classroom and school level (Table 4.2). Student performance on 8^{th} grade end-of-grade exams in reading are the strongest predictor of English I test performance. In the base model, end-of-grade tests in both 8^{th} grade reading and math indicate a statistically significant linear and non-linear relationship on English I test score outcomes. Figure 4.1 provides a graphical representation of the predicted impact of 8^{th} grade reading test performance over the relevant range of test scores on the English I end-of-grade test. The non-linear terms on prior reading test performance show

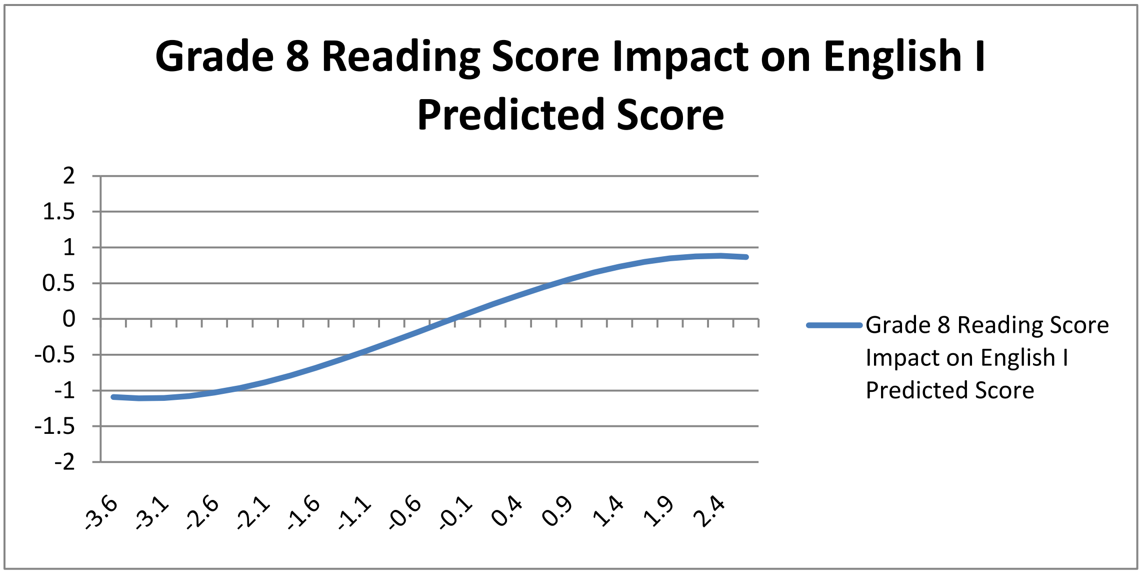

Figure 4.1 – Grade 8 Reading Score Impact on English I Predicted Score

that at about 2.4 standard deviations above the mean and about 3.1 standard deviations below the mean, the predicted impact of prior test scores is maximized or minimized respectively. While the coefficients on these three terms, Std Read Score (Grade 8), Std Read Score Squared (Grade 8), and Std Read Score Cubed (Grade 8), vary slightly across models, their statistical significance is maintained.

Figure 4.2 shows the relationship between prior test performance in 8th grade math and English I test performance. Only the Std Math Score (Grade 8) and Std Math Score Squared (Grade 8) were statistically significant across all of the peer effects related models in English I. As prior math performance increases, the influence on English I test performance is increasing at a slightly increasing rate. At the highest value of student prescores, about 2.9 standard deviations above zero, the student is expected to score about 0.72 standard deviations higher on the English I exam compared to a similar student who scored at the mean (0) on their 8th grade math exam.

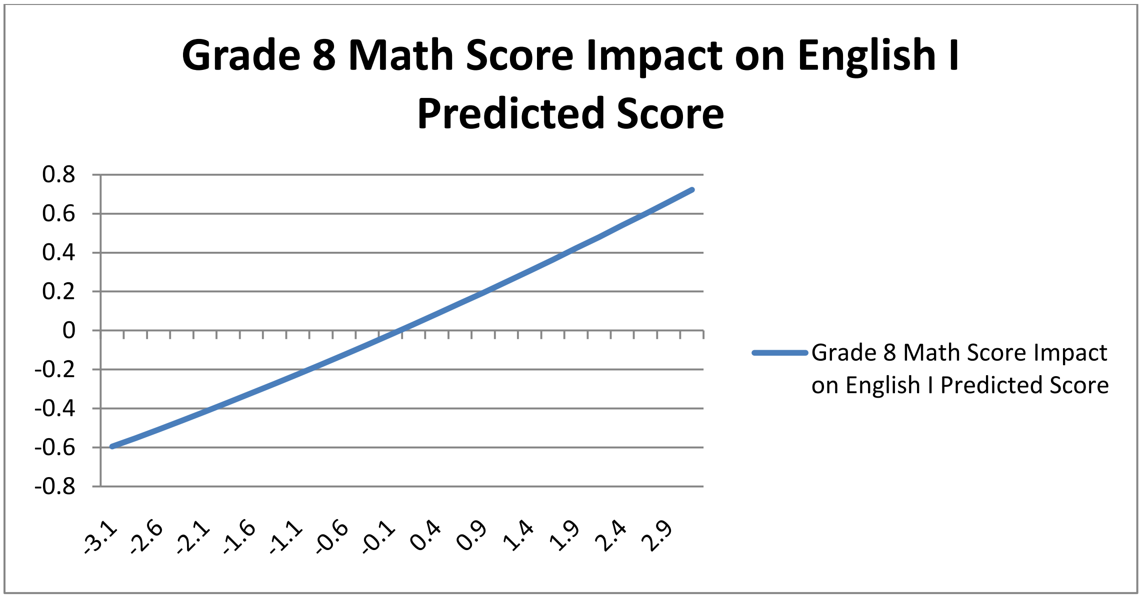

Figure 4.2 –Grade 8 Math Score Impact on English I Predicted Score

Male students perform more poorly than similar female students, scoring about 0.15 standard deviations lower than comparable female students. White and Hispanic ethnicity students perform similarly on the English I exam, while Black students and students of race other perform worse (0.08 and 0.02 standard deviations lower respectively). Underage students score higher (0.07 standard deviations) and overage students perform lower (about 0.09 standard deviations) than similar students who are aged within the cutoffs for their grade. Students coded as academically or intellectually gifted perform, on average, better than similar students (about 0.12 standard deviations), while those coded as disabled perform worse than non-exceptionally coded students (about ¼ of a standard deviation).

English I test takers coded as receiving reduced lunches score as well as those paying full price for lunch, but free lunch receipt students score slightly lower on average, by about 0.04 standard deviations. Parental education also predicts test performance at statistically significant levels. Compared to students with parents who are high school graduates, students with parents who did not graduate high school or those whose parent education measure is missing score lower (0.03 and 1/10 of a standard deviation lower respectively). Students whose parents attended either some college (0.06 standard deviations) or graduated from college (0.07 standard deviations) score higher than similar students whose parents are high school graduates only.

Students receiving services as Limited English Proficiency (LEP) students (0.14 standard deviations lower) and those who formerly received LEP services (0.05 standard deviations lower) score lower than similar students who never received LEP services. Compared to 9th grade English I students, students in higher grades perform better on the English I EOC exam with students in the highest grades performing best. On average and compared to 9th grade students, 12th grade students score about 0.32 standard deviations higher, 11th grade students score about 0.23 standard deviations higher, and 10th grade students score about 0.06 standard deviations higher compared to similar students. As expected, missing school days is associated with decreased performance on the EOC English I exam. Each additional day of school missed decreases a student's predicted exam score by about 0.005 standard deviation units compared to a similar student with one fewer days absent. A student missing 10 days is expected to score about 0.05 standard deviations lower than a similar student who did not miss any school days.

The base model's classroom characteristic variables reveal some differences in predicted student outcomes as a result of differing conditions across classrooms and the teachers within classrooms (Table 4.2 Model (1)). Compared to a class size with the mean number of students, larger classrooms on average perform slightly better than the average size classroom. Increasing the size of the class by one student above the average class size, is expected to increase student outcomes by about 0.005 standard deviations compared to students in similar classrooms. This counterintuitive finding may be explained by possible endogeneity between class size and student achievement. Schools could place students in smaller classes when they expect poor performance and use small classes as a mechanism for improving the performance of typically low performing students (Angrist & Lavy, 1999).

The ethnic configuration of classrooms only influences student outcomes in the base model (Table 4.2 Model (1)), where the ability level of peers from the prior year is omitted. In the base model, where prior academic performance of a student's classmates is ignored, classrooms with higher percentages of Hispanic students perform more poorly than similar classrooms with a lower percentage of students from this ethnic group (0.002 standard deviations). These differences are quite small, however, and a 10 percentage point increase in the proportion of Hispanic students within a classroom from the mean classroom would result in an expected decrease of 0.02 standard deviations in test performance for each student within the class. This suggests that for English I students, the skill level of peers is more important than a classroom's ethnic or income configuration.

Enrollment in classrooms providing an advanced curriculum is associated with significantly better test score performance (0.16 standard deviations higher on average) in Model (1) compared to similar students in classrooms provided with the regular curriculum (neither advanced nor remedial) (Table 4.2). The estimated impact of curriculum is diminished once the ability level of a student's peers is taken into account.

The predicted impact of an advanced curriculum class decreases by half to about 0.08 standard deviations in models including both linear and non-linear peer ability variables (Table 4.2 Model (3)). Students enrolled in remedial curriculum English I classrooms perform no differently on average than those in regular instruction classrooms.

Teacher characteristics impacting student outcomes in the base model of English I performance include the teacher's licensure type, infield status, and years of experience. The students of teachers holding licenses coded as emergency, temporary, and provisional performed slightly better on average than similar students whose instructors held initial or continuing licenses (0.04 standard deviations). Students of infield teachers, those with continuing or initial licenses in High School English, performed better (0.03 standard deviations) than similar students in classrooms taught by an instructor whose credential was in a different subject or grade level. Students in classrooms whose teachers were in their first year of teaching scored worse (0.03 standard deviations) and those with teachers having more than 25 years of experience performed better (0.03 standard deviations) on average, compared to similar students whose teachers had levels of experience between these two extremes. Once additional variables, including Peer Ability, are included in the model, no teacher characteristics remain statistically significant in their impact on student outcomes. This suggests that teacher/student sorting is responsible for the observed relationships in the previous model.

These models also include some school characteristics that appear to influence the English I test performance of students. Comparing similar students and classrooms in different school settings, students within schools containing a higher than average percentage of Hispanic students appear to perform better than students in schools with an average number of Hispanic students. A student in a school whose percentage of Hispanic students is one percent higher than the average school would be expected to score about 0.004 standard deviations higher in English I than a similar student. Increasing concentrations of poverty within schools has the opposite effect. Comparable students in a school whose percent of students receiving free lunch was 10 percent higher than average is associated with a score on average about 0.03 standard deviations lower on the English I EOC exam.

Algebra I Findings

The first step in modeling began with consideration of a null model designed to provide the level of variance in Algebra I test score outcomes by considering how variation in the outcome variable is partitioned across each of the three levels in the HLM model.

Table 4.3 - Algebra I – ICC Calculation

	Variance	Intraclass Correlation Coefficient
Level I - Student	0.468	52.5%
Level II - Classroom	0.206	23.2%
Level III - School	0.216	24.3%

Table 4.3 indicates the intraclass correlation coefficients for each level in the Algebra I null model. While variation in Algebra I scores is more evenly split between classroom and school levels, the majority of outcome score variation is between students within classrooms. The school level actually accounts for a slightly higher proportion of variance than classrooms (24.3 percent versus 23.2 percent). As expected, the intraclass

correlation coefficients indicate that most variation occurs at the student level with significant additional variation at the classroom level.

Table 4.4 - Algebra I Results

		(1)		(2)		(3)	
		Base Model - A1 No Peer Effects		Model - A1 With Peer Effects - Reading		Model - A1 With Peer Effects - Reading Squared and Cubed	
Coefficient Group		Coef.	Std. Err	Coef.	Std. Err	Coef.	Std. Err
School-Level	Intercept	-0.012	(0.034)	0.025	(0.032)	0.024	(0.032)
	Tracking Quartile - Lowest	----	----	----	----	----	----
	Tracking Quartile - Medium	----	----	----	----	----	----
	Tracking Quartile - Highest	----	----	----	----	----	----
	Total Per Pupil Exp (excl. Capital & Com. Ser.)	-0.003	(0.001)	-0.003	(0.001)	-0.003	(0.001)
	Average Daily Membership	0.000	(0.000)	0.000	(0.000)*	0.000	(0.000)*
	Teacher Turnover	-0.006	(0.002)**	-0.005	(0.002)*	-0.005	(0.002)*
	School Pct Black	-0.001	(0.001)	0.000	(0.001)	0.000	(0.001)
	School Pct Hispanic	-0.001	(0.003)	0.000	(0.003)	0.000	(0.003)
	School Pct Other	-0.002	(0.002)	-0.001	(0.002)	-0.001	(0.002)
	School Pct Free Lunch	-0.002	(0.001)	-0.001	(0.001)	-0.001	(0.001)
	School Pct Reduced Lunch	0.011	(0.006)	0.008	(0.005)	0.008	(0.005)
	Urban Area School	-0.068	(0.046)	-0.057	(0.043)	-0.058	(0.043)
	Rural Area School	-0.022	(0.033)	-0.009	(0.031)	-0.010	(0.031)

Table 4.4 - Algebra I Results – Continued

Classroom-Level							
	Peer Dispersion - Math (Std. Dev.)	----	----	----	----	----	----
	Number of Students in Class	0.004	(0.001)**	0.001	(0.001)	0.001	(0.001)
	Classroom Percent Black	-0.002	(0.000)**	0.000	(0.000)	0.000	(0.000)
	Classroom Percent Hispanic	-0.001	(0.000)	0.001	(0.000)	0.000	(0.000)
	Classroom Percent Other	-0.001	(0.001)	-0.001	(0.000)	-0.001	(0.000)
	Classroom Percent Free Lunch Eligible	-0.002	(0.000)**	-0.001	(0.000)	-0.001	(0.000)
	Classroom Percent Reduced Lunch Eligible	-0.001	(0.000)	0.000	(0.000)	0.000	(0.000)
	Curriculum Advanced	0.030	(0.026)	0.035	(0.023)	0.028	(0.023)
	Curriculum Remedial	0.003	(0.026)	-0.009	(0.024)	-0.011	(0.024)
	Advanced Degree	-0.019	(0.015)	-0.024	(0.015)	-0.023	(0.015)
	Licensed Other	-0.055	(0.027)*	-0.056	(0.027)*	-0.057	(0.027)*
	National Board Certification	0.079	(0.021)**	0.083	(0.021)**	0.083	(0.021)**
	Infield Teacher	0.018	(0.017)	0.008	(0.016)	0.008	(0.016)
	Most or Highly Competitive Barron's Rating	0.030	(0.017)	0.027	(0.017)	0.026	(0.017)
	First Year Teacher	-0.084	(0.029)**	-0.078	(0.028)**	-0.079	(0.028)**
	1 to 2 Years Teacher Experience	-0.032	(0.026)	-0.028	(0.026)	-0.027	(0.026)
	3 to 4 Years Teacher Experience	0.019	(0.030)	0.019	(0.029)	0.019	(0.029)
	10 to 14 Years Teacher Experience	-0.004	(0.023)	0.000	(0.023)	-0.001	(0.023)
	15 to 19 Years Teacher Experience	-0.005	(0.023)	-0.007	(0.023)	-0.009	(0.023)

Table 4.4 - Algebra I Results – Continued

	20 to 24 Years Teacher Experience	0.019	(0.029)	0.013	(0.027)	0.012	(0.026)
	25 or more Years Teacher Experience	-0.014	(0.022)	-0.016	(0.022)	-0.017	(0.022)
	Teacher Test Avg (Praxis, NTE, etc.)	-0.002	(0.011)	-0.003	(0.011)	-0.003	(0.011)
	Teacher Non-Certified Pay	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
Individual-Level							
	Peer Ability Read	----	----	----	----	----	----
	Peer Ability Math	----	----	0.135	(0.016)**	0.115	(0.019)**
	Peer Ability Math Squared	----	----	----	----	0.011	(0.016)
	Peer Ability Math Cubed	----	----	----	----	0.021	(0.012)
	Low-Performing Student * Peer Ability Math	----	----	----	----	----	----
	High-Performing Student * Peer Ability Math	----	----	----	----	----	----
	Std Math Score (Grade 8)	0.624	(0.010)**	0.621	(0.010)**	0.621	(0.010)**
	Std Math Score Squared (Grade 8)	0.079	(0.004)**	0.077	(0.004)**	0.077	(0.005)**
	Std Math Score Cubed (Grade 8)	-0.029	(0.003)**	-0.029	(0.003)**	-0.029	(0.003)**
	Std Read Score (Grade 8)	0.103	(0.006)**	0.102	(0.006)**	0.102	(0.006)**
	Std Read Score Squared (Grade 8)	-0.001	(0.003)	-0.001	(0.003)	-0.001	(0.003)
	Std Read Score Cubed (Grade 8)	-0.003	(0.001)	-0.003	(0.001)	-0.003	(0.001)
	Male	-0.031	(0.007)**	-0.031	(0.007)**	-0.030	(0.007)**
	Black	-0.078	(0.008)**	-0.081	(0.008)**	-0.081	(0.008)**
	Hispanic	-0.032	(0.018)	-0.034	(0.018)	-0.034	(0.018)

Table 4.4 - Algebra I Results – Continued

	Other	0.011	(0.014)	0.010	(0.014)	0.010	(0.014)
	Underage Student based on Grade	0.146	(0.028)**	0.147	(0.028)**	0.147	(0.028)**
	Overage Student based on Grade	-0.108	(0.007)**	-0.106	(0.007)**	-0.106	(0.007)**
	Academically or Intellectually Gifted	0.195	(0.014)**	0.190	(0.015)**	0.190	(0.015)**
	Disabled Student	-0.106	(0.012)**	-0.099	(0.012)**	-0.098	(0.012)**
	Free Lunch	0.011	(0.007)	0.009	(0.007)	0.009	(0.007)
	Reduced Lunch	0.003	(0.011)	0.003	(0.011)	0.003	(0.011)
	Parent Education Less than High School	0.017	(0.012)	0.017	(0.012)	0.017	(0.012)
	Parent Education Some College	0.021	(0.007)**	0.021	(0.007)**	0.021	(0.007)**
	Parent Education College Graduate	0.027	(0.008)**	0.026	(0.008)**	0.026	(0.008)**
	Parent Education Missing	-0.061	(0.038)	-0.064	(0.038)	-0.064	(0.038)
	Previous LEP Services Recipient	0.033	(0.030)	0.030	(0.030)	0.030	(0.030)
	LEP Services Recipient	0.041	(0.024)	0.039	(0.024)	0.039	(0.024)
	Grade 10	-0.013	(0.011)	-0.006	(0.011)	-0.006	(0.011)
	Grade 11	-0.019	(0.017)	-0.009	(0.017)	-0.010	(0.017)
	Grade 12	-0.041	(0.029)	-0.031	(0.029)	-0.031	(0.029)
	Days Absent	-0.009	(0.000)**	-0.009	(0.000)**	-0.009	(0.000)**
	Low-Performing Student	----	----	----	----	----	----
	High-Performing Student	----	----	----	----	----	----

Table 4.4 - Algebra I Results – Continued

		(4)		(5)		(6)	
		Model - A1 With Peer Effects - Reading Only - Peer Dispersion		Model - A1 With Peer Effects - Reading Only - Peer Dispersion - Low and High Performing Student Interactions		Model - A1 With Peer Effects - Reading Only - Tracking	
Coefficient Group		Coef.	Std. Err	Coef.	Std. Err	Coef.	Std. Err
School-Level	Intercept	0.055	(0.037)	0.022	(0.032)	-0.082	(0.038)*
	Tracking Quartile - Lowest	----	----	----	----	0.063	(0.035)
	Tracking Quartile - Medium	----	----	----	----	0.060	(0.039)
	Tracking Quartile - Highest	----	----	----	----	0.147	(0.047)**
	Total Per Pupil Exp (excl. Capital & Com. Ser.)	-0.003	(0.001)	-0.003	(0.001)	-0.003	(0.001)
	Average Daily Membership	0.000	(0.000)*	0.000	(0.000)*	0.000	(0.000)
	Teacher Turnover	-0.005	(0.002)*	-0.005	(0.002)*	-0.006	(0.002)**
	School Pct Black	0.000	(0.001)	0.000	(0.001)	-0.001	(0.001)
	School Pct Hispanic	0.000	(0.003)	0.000	(0.003)	0.000	(0.003)
	School Pct Other	-0.001	(0.002)	-0.001	(0.002)	-0.002	(0.001)
	School Pct Free Lunch	-0.001	(0.001)	-0.001	(0.001)	-0.001	(0.001)
	School Pct Reduced Lunch	0.008	(0.005)	0.008	(0.005)	0.011	(0.005)
	Urban Area School	-0.055	(0.043)	-0.058	(0.043)	-0.057	(0.046)
	Rural Area School	-0.009	(0.031)	-0.010	(0.031)	-0.018	(0.033)

Table 4.4 - Algebra I Results – Continued

Classroom-Level							
	Peer Dispersion - Math (Std. Dev.)	-0.051	(0.029)	----	----	----	----
	Number of Students in Class	0.001	(0.001)	0.002	(0.001)	0.005	(0.001)**
	Classroom Percent Black	0.000	(0.000)	0.000	(0.000)	-0.002	(0.000)**
	Classroom Percent Hispanic	0.001	(0.000)	0.001	(0.000)	0.000	(0.000)
	Classroom Percent Other	-0.001	(0.000)	-0.001	(0.000)	-0.001	(0.001)
	Classroom Percent Free Lunch Eligible	-0.001	(0.000)	-0.001	(0.000)	-0.002	(0.000)**
	Classroom Percent Reduced Lunch Eligible	0.000	(0.000)	0.000	(0.000)	-0.001	(0.000)
	Curriculum Advanced	0.034	(0.023)	0.030	(0.023)	----	----
	Curriculum Remedial	-0.010	(0.024)	-0.012	(0.024)	----	----
	Advanced Degree	-0.023	(0.015)	-0.023	(0.015)	----	----
	Licensed Other	-0.057	(0.027)*	-0.057	(0.027)*	----	----
	National Board Certification	0.083	(0.021)**	0.083	(0.021)**	----	----
	Infield Teacher	0.008	(0.016)	0.009	(0.016)	----	----
	Most or Highly Competitive Barron's Rating	0.027	(0.017)	0.027	(0.017)	----	----
	First Year Teacher	-0.078	(0.028)**	-0.077	(0.028)**	----	----
	1 to 2 Years Teacher Experience	-0.027	(0.026)	-0.026	(0.026)	----	----
	3 to 4 Years Teacher Experience	0.020	(0.029)	0.019	(0.029)	----	----
	10 to 14 Years Teacher Experience	0.001	(0.023)	0.001	(0.023)	----	----

Table 4.4 - Algebra I Results – Continued

	15 to 19 Years Teacher Experience	-0.007	(0.023)	-0.008	(0.023)	----	----
	20 to 24 Years Teacher Experience	0.014	(0.027)	0.012	(0.027)	----	----
	25 or more Years Teacher Experience	-0.016	(0.022)	-0.015	(0.022)	----	----
	Teacher Test Avg (Praxis, NTE, etc.)	-0.003	0.0000	-0.003	(0.011)	----	----
	Teacher Non-Certified Pay	0.000	0.0000	0.000	(0.000)	----	----
Individual-Level							
	Peer Ability Read	----	----	----	----	----	----
	Peer Ability Math	0.135	(0.016)**	0.130	(0.018)**	----	----
	Peer Ability Math Squared	----	----	----	----	----	----
	Peer Ability Math Cubed	----	----	----	----	----	----
	Low-Performing Student * Peer Ability Math	----	----	-0.025	(0.019)	----	----
	High-Performing Student * Peer Ability Math	----	----	0.061	(0.021)**	----	----
	Std Math Score (Grade 8)	0.621	(0.010)**	0.628	(0.010)**	0.624	(0.010)**
	Std Math Score Squared (Grade 8)	0.077	(0.004)**	0.072	(0.005)**	0.079	(0.005)**
	Std Math Score Cubed (Grade 8)	-0.029	(0.003)**	-0.030	(0.003)**	-0.029	(0.003)**
	Std Read Score (Grade 8)	0.102	(0.006)**	0.113	(0.008)**	0.103	(0.006)**
	Std Read Score Squared (Grade 8)	-0.001	(0.003)	-0.002	(0.004)	-0.001	(0.003)
	Std Read Score Cubed (Grade 8)	-0.003	(0.001)	-0.005	(0.001)*	-0.003	(0.001)
	Male	-0.030	(0.007)**	-0.030	(0.007)**	-0.032	(0.007)**
	Black	-0.081	(0.008)**	-0.082	(0.008)**	-0.078	(0.008)**

Table 4.4 - Algebra I Results – Continued

	Hispanic	-0.034	(0.018)	-0.034	(0.018)	-0.031	(0.018)
	Other	0.010	(0.014)	0.009	(0.014)	0.011	(0.014)
	Underage Student based on Grade	0.147	(0.028)**	0.147	(0.028)**	0.146	(0.028)**
	Overage Student based on Grade	-0.106	(0.007)**	-0.106	(0.007)**	-0.107	(0.007)**
	Academically or Intellectually Gifted	0.190	(0.015)**	0.183	(0.015)**	0.195	(0.015)**
	Disabled Student	-0.098	(0.012)**	-0.100	(0.012)**	-0.106	(0.012)**
	Free Lunch	0.009	(0.007)	0.009	(0.007)	0.011	(0.007)
	Reduced Lunch	0.003	(0.011)	0.003	(0.011)	0.003	(0.011)
	Parent Education Less than High School	0.017	(0.012)	0.017	(0.012)	0.017	(0.012)
	Parent Education Some College	0.021	(0.007)**	0.021	(0.007)**	0.021	(0.007)**
	Parent Education College Graduate	0.026	(0.008)**	0.026	(0.008)**	0.027	(0.008)**
	Parent Education Missing	-0.064	(0.038)	-0.064	(0.038)	-0.061	(0.039)
	Previous LEP Services Recipient	0.030	(0.030)	0.032	(0.030)	0.033	(0.030)
	LEP Services Recipient	0.039	(0.024)	0.040	(0.024)	0.040	(0.024)
	Grade 10	-0.005	(0.011)	-0.005	(0.011)	-0.013	(0.011)
	Grade 11	-0.009	(0.017)	-0.009	(0.017)	-0.018	(0.017)
	Grade 12	-0.031	(0.029)	-0.030	(0.029)	-0.040	(0.029)
	Days Absent	-0.009	(0.000)**	-0.009	(0.000)**	-0.009	(0.000)**
	Low-Performing Student	----	----	0.018	(0.012)	----	----
	High-Performing Student	----	----	-0.018	(0.013)	----	----

Table 4.4 - Algebra I Results – Continued

		(7)		(8)		(9)	
		Model - A1 With Peer Effects - Reading Only - Tracking - Peer Ability Mediator		Model - A1 With Peer Effects - Reading Only - Tracking - Curriculum Mediator		Model - A1 With Peer Effects - Reading Only - Tracking - Teacher Characteristics Mediator	
Coefficient Group		Coef.	Std. Err	Coef.	Std. Err	Coef.	Std. Err
School-Level	Intercept	-0.043	(0.036)	-0.086	(0.039)*	-0.080	(0.044)
	Tracking Quartile - Lowest	0.049	(0.034)	0.062	(0.035)	0.061	(0.036)
	Tracking Quartile - Medium	0.043	(0.037)	0.058	(0.039)	0.061	(0.039)
	Tracking Quartile - Highest	0.121	(0.044)**	0.147	(0.046)**	0.145	(0.047)**
	Total Per Pupil Exp (excl. Capital & Com. Ser.)	-0.003	(0.001)	-0.003	(0.001)	-0.003	(0.001)
	Average Daily Membership	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
	Teacher Turnover	-0.005	(0.002)*	-0.006	(0.002)**	-0.005	(0.002)*
	School Pct Black	0.000	(0.001)	-0.001	(0.001)	-0.001	(0.001)
	School Pct Hispanic	0.000	(0.003)	0.000	(0.003)	0.000	(0.003)
	School Pct Other	-0.002	(0.001)	-0.002	(0.001)	-0.002	(0.002)
	School Pct Free Lunch	0.000	(0.001)	-0.001	(0.001)	0.000	(0.001)
	School Pct Reduced Lunch	0.009	(0.005)	0.012	(0.006)*	0.011	(0.005)
	Urban Area School	-0.049	(0.043)	-0.053	(0.046)	-0.052	(0.045)
	Rural Area School	-0.006	(0.030)	-0.019	(0.033)	-0.021	(0.032)

Table 4.4 - Algebra I Results – Continued

Classroom-Level							
	Peer Dispersion - Math (Std. Dev.)	----	----	----	----	----	----
	Number of Students in Class	0.001	(0.001)	0.005	(0.001)**	0.004	(0.001)**
	Classroom Percent Black	0.000	(0.000)	-0.002	(0.000)**	-0.002	(0.000)**
	Classroom Percent Hispanic	0.001	(0.000)	-0.001	(0.000)	-0.001	(0.000)
	Classroom Percent Other	-0.001	(0.000)	-0.001	(0.001)	-0.001	(0.001)
	Classroom Percent Free Lunch Eligible	-0.001	(0.000)	-0.002	(0.000)**	-0.002	(0.000)**
	Classroom Percent Reduced Lunch Eligible	-0.001	(0.000)	-0.001	(0.000)	-0.001	(0.000)
	Curriculum Advanced	----	----	0.030	(0.027)	----	----
	Curriculum Remedial	----	----	0.003	(0.027)	----	----
	Advanced Degree	----	----	----	----	-0.018	(0.015)
	Licensed Other	----	----	----	----	-0.054	(0.027)*
	National Board Certification	----	----	----	----	0.079	(0.021)**
	Infield Teacher	----	----	----	----	0.018	(0.017)
	Most or Highly Competitive Barron's Rating	----	----	----	----	0.030	(0.017)
	First Year Teacher	----	----	----	----	-0.084	(0.029)**
	1 to 2 Years Teacher Experience	----	----	----	----	-0.032	(0.026)
	3 to 4 Years Teacher Experience	----	----	----	----	0.019	(0.030)
	10 to 14 Years Teacher Experience	----	----	----	----	-0.004	(0.023)

Table 4.4 - Algebra I Results – Continued

	15 to 19 Years Teacher Experience	----	----	----	----	-0.006	(0.023)
	20 to 24 Years Teacher Experience	----	----	----	----	0.017	(0.029)
	25 or more Years Teacher Experience	----	----	----	----	-0.014	(0.022)
	Teacher Test Avg (Praxis, NTE, etc.)	----	----	----	----	-0.003	(0.011)
	Teacher Non-Certified Pay	----	----	----	----	0.000	(0.000)
Individual-Level							
	Peer Ability Read	----	----	----	----	----	----
	Peer Ability Math	0.135	(0.017)**	----	----	----	----
	Peer Ability Math Squared	----	----	----	----	----	----
	Peer Ability Math Cubed	----	----	----	----	----	----
	Low-Performing Student * Peer Ability Math	----	----	----	----	----	----
	High-Performing Student * Peer Ability Math	----	----	----	----	----	----
	Std Math Score (Grade 8)	0.621	(0.010)**	0.624	(0.010)**	0.623	(0.010)**
	Std Math Score Squared (Grade 8)	0.077	(0.004)**	0.079	(0.004)**	0.079	(0.004)**
	Std Math Score Cubed (Grade 8)	-0.0288	(0.003)**	-0.0293	(0.003)**	-0.0292	(0.003)**
	Std Read Score (Grade 8)	0.102	(0.006)**	0.103	(0.006)**	0.103	(0.006)**
	Std Read Score Squared (Grade 8)	-0.001	(0.003)	-0.001	(0.003)	-0.001	(0.003)
	Std Read Score Cubed (Grade 8)	-0.003	(0.001)	-0.003	(0.001)	-0.003	(0.001)
	Male	-0.031	(0.007)**	-0.032	(0.007)**	-0.031	(0.007)**
	Black	-0.081	(0.008)**	-0.078	(0.008)**	-0.078	(0.008)**

Table 4.4 - Algebra I Results – Continued

	Hispanic	-0.033	(0.018)	-0.031	(0.018)	-0.032	(0.018)
	Other	0.010	(0.014)	0.011	(0.014)	0.011	(0.014)
	Underage Student based on Grade	0.147	(0.028)**	0.146	(0.028)**	0.147	(0.028)**
	Overage Student based on Grade	-0.106	(0.007)**	-0.108	(0.007)**	-0.108	(0.007)**
	Academically or Intellectually Gifted	0.190	(0.015)**	0.195	(0.015)**	0.195	(0.014)**
	Disabled Student	-0.099	(0.012)**	-0.106	(0.012)**	-0.106	(0.012)**
	Free Lunch	0.009	(0.007)	0.011	(0.007)	0.011	(0.007)
	Reduced Lunch	0.003	(0.011)	0.003	(0.011)	0.003	(0.011)
	Parent Education Less than High School	0.017	(0.012)	0.017	(0.012)	0.017	(0.012)
	Parent Education Some College	0.021	(0.007)**	0.021	(0.007)**	0.021	(0.007)**
	Parent Education College Graduate	0.026	(0.008)**	0.027	(0.008)**	0.027	(0.008)**
	Parent Education Missing	-0.064	(0.039)	-0.061	(0.039)	-0.061	(0.038)
	Previous LEP Services Recipient	0.030	(0.030)	0.033	(0.030)	0.033	(0.030)
	LEP Services Recipient	0.038	(0.024)	0.040	(0.024)	0.041	(0.024)
	Grade 10	-0.005	(0.011)	-0.013	(0.011)	-0.013	(0.011)
	Grade 11	-0.008	(0.017)	-0.018	(0.017)	-0.018	(0.017)
	Grade 12	-0.030	(0.029)	-0.041	(0.029)	-0.040	(0.029)
	Days Absent	-0.009	(0.000)**	-0.009	(0.000)**	-0.009	(0.000)**
	Low-Performing Student	----	----	----	----	----	----
	High-Performing Student	----	----	----	----	----	----

Peer Ability and Algebra I Achievement

Due to the strong predictive power of prior math performance on Algebra I test outcomes, I use prior math achievement to measure peer prior achievement for Algebra I models. This series of models for Algebra I also differs from English I and Biology models in that models for Algebra I implement the inverse mills ratio weights discussed in Appendix A.

Predicting that students with higher performing peers (measured by averaging the 8^{th} grade math test performance of all of the other students in a classroom) will score better on the Algebra I exam compared to similar students in classrooms with lower performing peers forms the basis for Research Question 1. The mechanism for this process might be through cross-student learning or a more quickly paced curriculum that allows these classrooms to cover more ground than classrooms with lower performing students. Research Question 1 is supported in the model results 'A1 with Peer Effects – Math Only' (Table 4.4 Model (2)). Placing a student in a class where one's peers scored one standard deviation higher than average is associated with an increase in that student's predicted performance on the end-of-course Algebra I exam by about 0.135 standard deviations. If the reverse is true and students are placed in classrooms with lower performing peers, they are expected to perform similarly worse than comparable students. As observed in the previous models focused on English I achievement, I test for a non-linear relationship between Algebra I test score achievement and prior peer performance on eighth grade math test performance. Model (3) tests for non-linear impacts of peer ability on a student's Algebra I test performance (Table 4.4). As a result of this analysis, I reject the idea that the influence of peer ability on Algebra I test performance is non-linear. The models 'Peer Ability Math Squared' and 'Peer Ability Math Cubed' both have statistically insignificant coefficients (Table 4.4 Model (3)). Prior math peer achievement is associated with constant returns over the relevant range of peer ability values.

Peer Dispersion and Algebra I Achievement

A mathematics classroom may provide significant opportunities for cross-peer learning and group work as students of similar or quite different ability may be grouped to tackle assignments in ways that might be more difficult in a reading or literature classroom. Research Question 2 predicts that students in more academically diverse classrooms (those where the dispersion of prior math test performance is greater) will score better on Algebra I exams compared to similar students. The results of Model (4) reject Research Question 2 as there is no statistically significant difference in test performance between similar students in classrooms that are more or less academically dispersed (Peer Dispersion - Math (Std. Dev.) in Table 4.4).

The final model, focused on peer differences, tests whether the impact of peer ability on student test performance differs for students classified as high or low performing students (top and bottom 20 percent of students) in either reading or math based on 8^{th} grade test performance. In Model (5) for Algebra I, I find that high performing students differentially benefit from being in a class with high performing peers (Table 4.4). The coefficient on the term which interacts the value of peer ability with high performing students suggests that while the overall returns to the average student from being in a classroom with peers who score 1 standard deviation above average on their 8th grade math exam compared to a student with average performing peers is about 0.13 standard deviations, but high performing students score an additional 0.061 standard deviations higher for a total benefit of about 0.21 standard deviations (0.13 + 0.061). The returns to higher peer ability are larger for high performing students but would also differentially diminish their scores when placed in a classroom with peers who scored lower than average on their 8th grade math exam. For low performing students, I reject the hypothesis that the returns to these students from peer effects is different from zero as the coefficient on Low Performing Student * Peer Ability – Math is not significantly different from zero (Table 4.4 Model (5)). This model suggests that there is no additional relationship between peer math ability and EOC test performance for low performing students, but that high performing students in classrooms with high levels of peer ability score better than would be expected based on their peer's math ability alone.

Tracking and Algebra I Achievement

In Algebra I schools with higher levels of tracking, I find some support for Research Question 3. Schools with higher intensity tracking have higher student outcomes on EOC tests at statistically significant levels. Algebra I students enrolled in schools with the highest levels of tracking scored significantly higher (about 0.147 standard deviations) than students in schools with lower levels of tracking or schools with no evidence of tracking (Table 4.4 Model (6)). Algebra I student scores in schools with lower levels of tracking were not statistically different than scores for students enrolled in un-tracked schools. Subsequent models tested whether three suspected mediators (peer ability, curriculum differences, or teacher characteristics) accounted for the difference in Algebra I EOC test score outcomes for students in the most highly tracked schools and students in other schools. All of the Algebra I tracking models continue to use the inverse mills ratio weights previously discussed in the data section.

The first mediator model, which includes a measure of a student's classroom peers based on their grade 8 math test achievement, reduces the coefficient on the highest quartile of tracked schools to 0.122 standard deviations from the 0.147 SD effect observed in the initial model (Table 4.4 Model (7)). This is a decrease of about 17 percent ((0.147 – 0.122) / 0.147) in the magnitude of the coefficient. I conclude that at least some of the associated increase in test score performance between students in high intensity tracked schools is due to their placement in classrooms with higher performing peers. However, this mediator does not fully explain the association observed for students in highly tracked schools. The final two mediators, curriculum differences and teachers characteristics, fail to substantially reduce the coefficient on the most highly tracked quartile school students (Table 4.4 Columns (8) and (9)). This result indicates that the measured curriculum differences and teacher characteristics are not possible explanations for the observed positive effects of high intensity tracked schools.

Algebra I Control Variables

The base model for Algebra I students reinforces the strong influence of individual characteristics on student outcomes and lesser impacts of characteristics measured at the classroom and school level similarly observed in the prior subject models (Table 4.4 Model (1)). A student scoring one standard deviation higher than average on the 8th grade reading test predicts an Algebra I test score 0.103 standard deviations higher than average while an 8th grade math test score that is one standard deviation higher than average predicts an Algebra I test score that is 0.624 standard deviations above average, holding other characteristics constant (Table 4.4 Model (2)). There is little variation observed in the calculated values of coefficients across Algebra I models, and the remaining values described in this section are based on the values in Model (1) (Table 4.4). Student performance on 8th grade end-of-grade exams in math are the strongest predictor of Algebra I test performance. In the base model (Column 1), end-of-grade tests in both reading and math had statistically significant linear effects, while prior math performance also revealed statistically significant non-linear coefficients (Table 4.4).

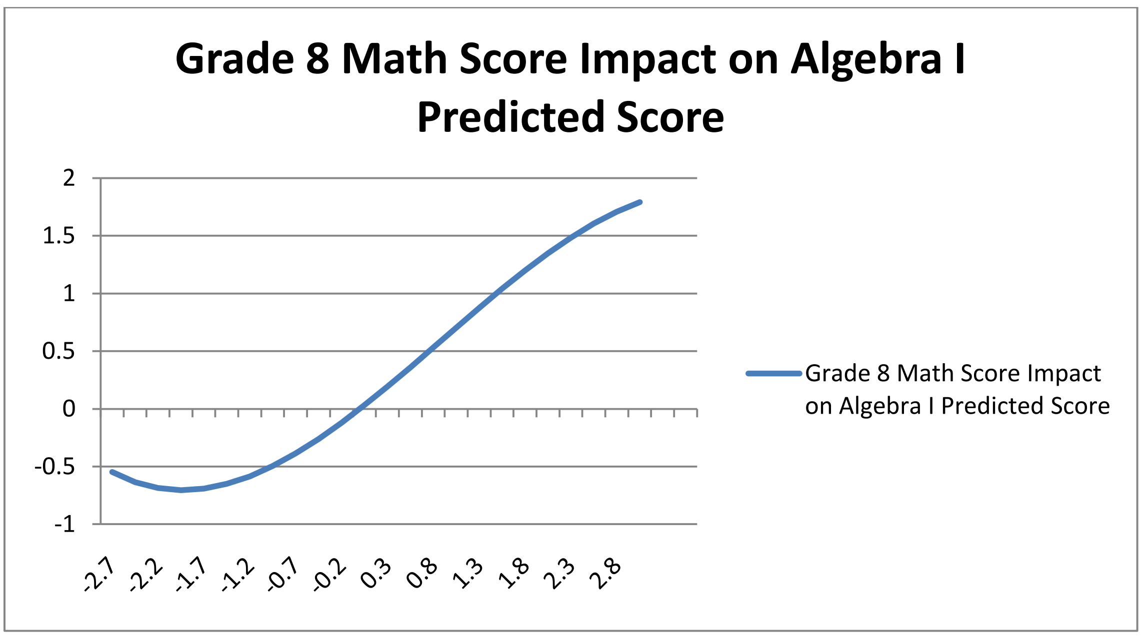

Figure 4.3 – Grade 8 Math Score Impact on Algebra I Predicted Score

Figure 4.3 provides a graphical representation of the impact of 8th grade math test performance over the relevant range of test scores on the Algebra I end-of-grade test. The non-linear terms on prior math test performance show that near the maximum values for prior math scores, the impact of prior math performance is increasing at a decreasing rate.

A student who scored 2.8 standard deviations above the mean on their 8th grade math exam would be expected to score about 1.71 standard deviations higher on the Algebra I exam compared to a similar student scoring at the mean on the prior exam. Students scoring 1.95 standard deviations below the mean on the 8th grade math exam are expected to score about 0.7 standard deviations below average compared to similar students scoring at the mean in 8th grade. This value is about the maximum negative impact observed in Figure 3. While the coefficients on these three terms, Std Math Score (Grade 8), Std Math Score Squared (Grade 8), and Std Math Score Cubed (Grade 8), vary slightly across models, their statistical significance is consistent.

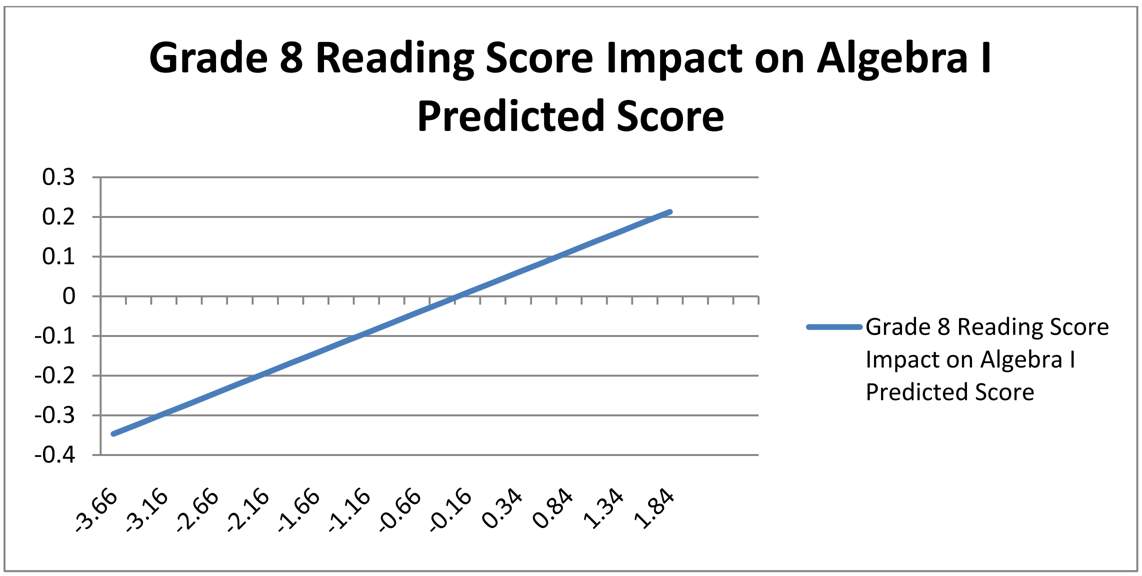

Figure 4.4 – Grade 8 Reading Score Impact on Algebra I Predicted Score

Figure 4.4 shows the expected impact of prior performance on 8th grade reading scores on Algebra I test performance. Only the Std Read Score (Grade 8) is statistically significant across all of the peer effects related models in Algebra I. As prior reading performance increases, the influence on Algebra I test performance increases at a constant rate over the range of relevant scores. At the highest value of student prescores, about 2.09 standard deviations above zero, the student is expected to score about 0.21 standard deviations higher on the English I exam compared to a similar student who scored at the mean (0) on their 8th grade math exam.

Male students perform slightly more poorly than similar female students, scoring about 0.03 standard deviations lower than comparable female students. Students with Black ethnicity perform more poorly on average than students in other ethnic categories, about 0.08 standard deviations lower on average when compared to similar students who only differ based on ethnicity. Underage students score substantially higher (0.14 standard deviations) and overage students perform lower (about 1/10 of a standard deviation) than similar students who are aged within the cutoffs for their grade. Students coded as academically or intellectually gifted perform, on average, better than similar students (about 0.19 standard deviations), while those coded as disabled perform worse than non-exceptionally coded students (about 0.10 standard deviations) (Table 4.4 Model (1)).

Algebra I test takers receiving free or reduced priced lunches scored as well as the full pay reference group. Parental education also predicts test performance at statistically significant levels. Compared to students with parents who did not finish high school or are high school graduates, students whose parents attended either some college or graduated from college score about 0.02 standard deviations higher than similar students (Table 4.4 Model (1)).

Students receiving services as Limited English Proficiency students and those who formerly received LEP services score about as well on the Algebra I exam as similar students who never received LEP services. Students similar on other characteristics perform about the same in Algebra I, regardless of grade level. As expected, missing school days leads to decreased performance on the EOC Algebra I exam, each additional day of school missed decreases a student's exam score by about 0.009 standard deviation units compared to a similar student with one fewer days absent. A student missing 10 days would be expected to score about 0.09 standard deviations lower than a similar student who did not miss any school days (Table 4.4 Model (1)).

The base model's classroom characteristic variables reveal some differences in student outcomes as a result of differing conditions across classrooms and the teachers within classrooms (Table 4.4 Model (1)). Compared to a class size with the mean number of students, larger classrooms on average perform slightly better than the average size classroom only in the model that excludes the influence of peer ability. In the other models, class size is not a significant predictor of student performance.

The ethnic and income configuration of classrooms only influences student outcomes in the base model, where the ability level of peers from the prior year is not considered (Table 4.4 Model (1)). In the base model, where prior academic performance of a student's classmates is ignored, classrooms with higher percentages of black students perform more poorly than similar classrooms with a lower percentage of students from this ethnic group (0.002 standard deviations) (Table 4.4 Model (1)). Classrooms with higher than average percentages of free lunch students also predicts lower test performance in the base model (0.002 standard deviations). Similar to the predicted outcomes for English I students, the skill level of peers is more important than a classroom's ethnic or income configuration in Algebra I as well. There are no statistically significant differences in outcome for students enrolled in advanced or remedial Algebra I courses compared to similar students in regular curriculum courses.

Teacher characteristics impacting student outcomes were similar across Algebra I models. The teacher's licensure type, National Board Certification status, and years of experience were all significant predictors of student test performance in Algebra I classrooms. The students of teachers holding licenses coded as emergency, temporary, and provisional performed worse on average than similar students whose instructors held initial or continuing licenses (0.05 standard deviations) (Table 4.4 Model (1)). Students of teachers holding National Board Certification scored substantially higher (0.08 standard deviations) than similar students in classrooms taught by an instructor not Nationally Board Certified. Students in classrooms whose teachers were in their first year of teaching scored worse (about 0.08 standard deviations) on average compared similar students whose teachers had higher levels of experience (Table 4.4 Model (1)).

These models also include school characteristics that might influence the Algebra I test performance of students. Comparing similar students and classrooms in different school settings, only students within schools with higher teacher turnover had a statistically significant and non-zero impact on student performance. The relationship between school size and student performance was statistically significant but essentially zero (less than 0.000). A student in a school where teacher turnover was one percent higher than average scored about 0.005 standard deviations lower than comparable students in a school where teacher turnover was average. No other school characteristics had a statistically significant association with student test performance in Algebra I (Table 4.4 Model (1)).

Biology Findings

Table 4.5 presents the results of a null model designed to provide the level of variance in test score outcomes across each of the three levels in the HLM model.

Table 4.5 - Biology – ICC Calculation

	Variance	Intraclass Correlation Coefficient
Level I - Student	0.572	59.9%
Level II - Classroom	0.296	31.0%
Level III - School	0.087	9.1%

The results of the ICC calculations using Biology data mimic closely the results from the initial analysis conducted in English I. The ICC values decrease across each of the three levels with the highest variance occurring between students within classrooms and the lowest levels of variation in test score outcomes occur between schools.

Table 4.6 - Biology Results

		(1)		(2)		(3)	
		Base Model - Bio No Peer Effects		Model - Bio With Peer Effects - Read Only		Model - Bio With Peer Effects - NonLinear Test - Read Only	
Coefficient Group		Coef.	Std. Err	Coef.	Std. Err	Coef.	Std. Err
School-Level	Intercept	-0.318	(0.048)**	-0.289	(0.047)**	-0.305	(0.048)**
	Tracking Quartile - Lowest	----	----	----	----	----	----
	Tracking Quartile - Medium	----	----	----	----	----	----
	Tracking Quartile - Highest	----	----	----	----	----	----
	Total Per Pupil Exp (excl. Capital & Com. Ser.)	-0.002	(0.000)*	-0.002	(0.000)*	-0.002	(0.000)*
	Average Daily Membership	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
	Teacher Turnover	0.001	(0.001)	0.001	(0.001)	0.001	(0.001)
	School Pct Black	0.001	(0.001)	0.001	(0.001)	0.001	(0.001)
	School Pct Hispanic	0.002	(0.002)	0.002	(0.002)	0.002	(0.002)
	School Pct Other	0.002	(0.002)	0.002	(0.002)	0.002	(0.002)
	School Pct Free Lunch	-0.002	(0.002)	-0.001	(0.002)	-0.001	(0.002)
	School Pct Reduced Lunch	-0.002	(0.006)	-0.003	(0.005)	-0.003	(0.006)
	Urban Area School	-0.072	(0.043)	-0.071	(0.042)	-0.072	(0.043)
	Rural Area School	0.000	(0.030)	0.003	(0.030)	0.005	(0.030)

Table 4.6 - Biology Results – Continued

Classroom-Level							
	Peer Dispersion - Read (Std. Dev.)	----	----	----	----	----	----
	Number of Students in Class	-0.002	(0.001)	-0.003	(0.001)*	-0.003	(0.001)*
	Classroom Percent Black	-0.001	(0.000)	0.000	(0.000)	0.000	(0.000)
	Classroom Percent Hispanic	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
	Classroom Percent Other	0.000	(0.001)	0.001	(0.000)	0.001	(0.000)
	Classroom Percent Free Lunch Eligible	-0.001	(0.000)*	0.000	(0.000)	-0.001	(0.000)
	Classroom Percent Reduced Lunch Eligible	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
	Curriculum Advanced	0.218	(0.021)**	0.159	(0.023)**	0.136	(0.024)**
	Curriculum Remedial	0.040	(0.020)	0.041	(0.020)*	0.041	(0.020)*
	Advanced Degree	-0.011	(0.017)	-0.010	(0.017)	-0.009	(0.017)
	Licensed Other	0.038	(0.042)	0.023	(0.042)	0.032	(0.043)
	National Board Certification	0.051	(0.030)	0.048	(0.030)	0.047	(0.030)
	Infield Teacher	0.099	(0.035)**	0.078	(0.034)*	0.088	(0.036)*
	Most or Highly Competitive Barron's Rating	0.006	(0.022)	0.007	(0.022)	0.009	(0.022)
	First Year Teacher	-0.139	(0.036)**	-0.139	(0.036)**	-0.141	(0.036)**
	1 to 2 Years Teacher Experience	-0.022	(0.025)	-0.022	(0.025)	-0.023	(0.025)
	3 to 4 Years Teacher Experience	0.029	(0.030)	0.029	(0.030)	0.028	(0.030)
	10 to 14 Years Teacher Experience	-0.024	(0.033)	-0.028	(0.032)	-0.028	(0.032)
	15 to 19 Years Teacher Experience	-0.045	(0.031)	-0.048	(0.031)	-0.048	(0.030)
	20 to 24 Years Teacher Experience	0.025	(0.031)	0.022	(0.031)	0.022	(0.031)
	25 or more Years Teacher Experience	-0.078	(0.034)*	-0.084	(0.034)*	-0.083	(0.034)*
	Teacher Test Avg (Praxis, NTE, etc.)	0.018	(0.015)	0.016	(0.015)	0.016	(0.015)

Table 4.6 - Biology Results – Continued

	Teacher Non-Certified Pay	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
Individual-Level							
	Peer Ability Math	----	----	----	----	----	----
	Peer Ability Read	----	----	0.091	(0.017)**	0.098	(0.021)**
	Peer Ability Read Squared	----	----	----	----	0.033	(0.014)*
	Peer Ability Read Cubed	----	----	----	----	0.001	(0.011)
	Low-Performing Student * Peer Ability Read	----	----	----	----	----	----
	High-Performing Student * Peer Ability Read	----	----	----	----	----	----
	Std Math Score (Grade 8)	0.323	(0.008)**	0.322	(0.008)**	0.322	(0.008)**
	Std Math Score Squared (Grade 8)	0.041	(0.003)**	0.042	(0.003)**	0.041	(0.003)**
	Std Math Score Cubed (Grade 8)	-0.002	(0.001)	-0.003	(0.001)	-0.002	(0.001)
	Std Read Score (Grade 8)	0.429	(0.007)**	0.431	(0.007)**	0.431	(0.007)**
	Std Read Score Squared (Grade 8)	0.018	(0.003)**	0.018	(0.003)**	0.018	(0.003)**
	Std Read Score Cubed (Grade 8)	-0.020	(0.001)**	-0.020	(0.001)**	-0.020	(0.001)**
	Male	0.121	(0.006)**	0.122	(0.006)**	0.122	(0.006)**
	Black	-0.158	(0.008)**	-0.160	(0.008)**	-0.160	(0.008)**
	Hispanic	-0.019	(0.023)	-0.020	(0.023)	-0.021	(0.023)
	Other	-0.008	(0.014)	-0.009	(0.014)	-0.009	(0.014)
	Underage Student based on Grade	0.141	(0.027)**	0.141	(0.027)**	0.142	(0.027)**
	Overage Student based on Grade	-0.082	(0.008)**	-0.082	(0.008)**	-0.082	(0.008)**
	Academically or Intellectually Gifted	0.069	(0.013)**	0.068	(0.013)**	0.067	(0.013)**
	Disabled Student	-0.071	(0.015)**	-0.066	(0.015)**	-0.068	(0.015)**
	Free Lunch	0.011	(0.008)	0.009	(0.008)	0.009	(0.008)

Table 4.6 - Biology Results – Continued

	Reduced Lunch	0.038	(0.011)**	0.037	(0.011)**	0.037	(0.011)**
	Parent Education Less than High School	-0.014	(0.012)	-0.013	(0.012)	-0.013	(0.012)
Table 4.6 - Biology Results (cont)							
	Parent Education Some College	0.063	(0.007)**	0.062	(0.007)**	0.063	(0.007)**
	Parent Education College Graduate	0.076	(0.008)**	0.075	(0.009)**	0.075	(0.009)**
	Parent Education Missing	0.004	(0.073)	0.006	(0.073)	0.005	(0.073)
	Previous LEP Services Recipient	0.013	(0.038)	0.012	(0.038)	0.012	(0.038)
	LEP Services Recipient	-0.071	(0.030)*	-0.070	(0.030)*	-0.070	(0.030)*
	Grade 10	0.058	(0.012)**	0.059	(0.012)**	0.059	(0.012)**
	Grade 11	0.141	(0.018)**	0.143	(0.018)**	0.143	(0.018)**
	Grade 12	0.253	(0.027)**	0.254	(0.027)**	0.254	(0.027)**
	Days Absent	-0.007	(0.000)**	-0.007	(0.000)**	-0.007	(0.000)**
	Low-Performing Student	----	----	----	----	----	----
	High-Performing Student	----	----	----	----	----	----

Table 4.6 - Biology Results – Continued

		(4)		(5)		(6)	
		Model - Bio With Peer Effects - Math Only - Peer Dispersion		Model - Bio With Peer Effects - Math Only - Peer Dispersion - Low and High Performing Student Interactions		Model - Bio With Peer Effects - Math Only - Low Cutoff Tracking	
Coefficient Group		Coef.	Std. Err	Coef.	Std. Err	Coef.	Std. Err
School-Level	Intercept	-0.311	(0.051)**	-0.305	(0.048)**	-0.255	(0.042)**
	Tracking Quartile - Lowest	----	----	----	----	0.063	(0.033)
	Tracking Quartile - Medium	----	----	----	----	0.088	(0.036)*
	Tracking Quartile - Highest	----	----	----	----	0.124	(0.038)**
	Total Per Pupil Exp (excl. Capital & Com. Ser.)	-0.002	(0.000)*	-0.002	(0.000)*	-0.001	(0.000)
	Average Daily Membership	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
	Teacher Turnover	0.001	(0.001)	0.001	(0.001)	0.000	(0.002)
	School Pct Black	0.001	(0.001)	0.001	(0.001)	0.001	(0.001)
	School Pct Hispanic	0.002	(0.002)	0.002	(0.002)	0.004	(0.003)
	School Pct Other	0.002	(0.002)	0.002	(0.002)	0.003	(0.002)
	School Pct Free Lunch	-0.001	(0.002)	-0.001	(0.002)	-0.002	(0.002)
	School Pct Reduced Lunch	-0.003	(0.006)	-0.003	(0.006)	-0.006	(0.006)
	Urban Area School	-0.072	(0.043)	-0.072	(0.043)	-0.038	(0.043)
	Rural Area School	0.005	(0.030)	0.005	(0.030)	-0.003	(0.029)

Table 4.6 - Biology Results – Continued

Classroom-Level							
	Peer Dispersion - Read (Std. Dev.)	0.009	(0.031)	----	----	----	----
	Number of Students in Class	-0.003	(0.001)*	-0.003	(0.001)*	0.000	(0.001)
	Classroom Percent Black	0.000	(0.000)	0.000	(0.000)	-0.003	(0.000)**
	Classroom Percent Hispanic	0.000	(0.000)	0.000	(0.000)	-0.002	(0.000)*
	Classroom Percent Other	0.001	(0.000)	0.001	(0.000)	0.001	(0.001)
	Classroom Percent Free Lunch Eligible	0.000	(0.000)	0.000	(0.000)	-0.003	(0.000)**
	Classroom Percent Reduced Lunch Eligible	0.000	(0.000)	0.000	(0.000)	-0.002	(0.000)*
	Curriculum Advanced	0.136	(0.024)**	0.141	(0.025)**	----	----
	Curriculum Remedial	0.041	(0.020)*	0.042	(0.020)*	----	----
	Advanced Degree	-0.009	(0.017)	-0.009	(0.017)	----	----
	Licensed Other	0.032	(0.042)	0.030	(0.042)	----	----
	National Board Certification	0.047	(0.030)	0.047	(0.030)	----	----
	Infield Teacher	0.088	(0.036)*	0.086	(0.036)*	----	----
	Most or Highly Competitive Barron's Rating	0.008	(0.022)	0.008	(0.022)	----	----
	First Year Teacher	-0.141	(0.036)**	-0.140	(0.036)**	----	----
	1 to 2 Years Teacher Experience	-0.023	(0.025)	-0.022	(0.025)	----	----
	3 to 4 Years Teacher Experience	0.028	(0.030)	0.027	(0.030)	----	----
	10 to 14 Years Teacher Experience	-0.028	(0.032)	-0.028	(0.032)	----	----
	15 to 19 Years Teacher Experience	-0.048	(0.030)	-0.048	(0.031)	----	----

Table 4.6 - Biology Results – Continued

	20 to 24 Years Teacher Experience	0.022	(0.031)	0.022	(0.031)	----	----
	25 or more Years Teacher Experience	-0.083	(0.034)*	-0.084	(0.034)*	----	----
	Teacher Test Avg (Praxis, NTE, etc.)	0.017	(0.015)	0.016	(0.015)	----	----
	Teacher Non-Certified Pay	0.000	(0.000)	0.000	(0.000)	----	----
Individual-Level							
	Peer Ability Math	----	----	----	----	----	----
	Peer Ability Read	0.100	(0.017)**	0.082	(0.018)**	----	----
	Peer Ability Read Squared	0.033	(0.013)*	0.051	(0.014)**	----	----
	Peer Ability Read Cubed	----	----	----	----	----	----
	Low-Performing Student * Peer Ability Read	----	----	0.074	(0.029)*	----	----
	High-Performing Student * Peer Ability Read	----	----	-0.013	(0.017)	----	----
	Std Math Score (Grade 8)	0.322	(0.008)**	0.327	(0.009)**	0.330	(0.008)**
	Std Math Score Squared (Grade 8)	0.041	(0.003)**	0.044	(0.003)**	0.043	(0.003)**
	Std Math Score Cubed (Grade 8)	-0.002	(0.001)	-0.003	(0.001)	-0.003	(0.001)
	Std Read Score (Grade 8)	0.431	(0.007)**	0.435	(0.008)**	0.434	(0.007)**
	Std Read Score Squared (Grade 8)	0.018	(0.003)**	0.020	(0.003)**	0.019	(0.003)**
	Std Read Score Cubed (Grade 8)	-0.020	(0.001)**	-0.021	(0.001)**	-0.021	(0.001)**
	Male	0.122	(0.006)**	0.121	(0.006)**	0.119	(0.006)**
	Black	-0.160	(0.008)**	-0.160	(0.008)**	-0.154	(0.008)**
	Hispanic	-0.021	(0.023)	-0.021	(0.023)	-0.014	(0.023)

Table 4.6 - Biology Results – Continued

	Other	-0.009	(0.014)	-0.010	(0.014)	-0.007	(0.014)
	Underage Student based on Grade	0.142	(0.027)**	0.141	(0.027)**	0.143	(0.027)**
	Overage Student based on Grade	-0.082	(0.008)**	-0.081	(0.008)**	-0.085	(0.008)**
	Academically or Intellectually Gifted	0.067	(0.013)**	0.069	(0.014)**	0.085	(0.013)**
	Disabled Student	-0.068	(0.015)**	-0.067	(0.015)**	-0.072	(0.015)**
	Free Lunch	0.009	(0.008)	0.009	(0.008)	0.014	(0.008)
	Reduced Lunch	0.037	(0.011)**	0.037	(0.011)**	0.039	(0.011)**
	Parent Education Less than High School	-0.013	(0.012)	-0.013	(0.012)	-0.016	(0.012)
	Parent Education Some College	0.063	(0.007)**	0.062	(0.007)**	0.065	(0.007)**
	Parent Education College Graduate	0.075	(0.009)**	0.075	(0.009)**	0.080	(0.008)**
	Parent Education Missing	0.005	(0.073)	0.008	(0.072)	0.001	(0.074)
	Previous LEP Services Recipient	0.012	(0.038)	0.010	(0.038)	0.016	(0.038)
	LEP Services Recipient	-0.070	(0.030)*	-0.071	(0.031)*	-0.068	(0.030)*
	Grade 10	0.059	(0.012)**	0.058	(0.012)**	0.049	(0.012)**
	Grade 11	0.143	(0.018)**	0.143	(0.018)**	0.128	(0.018)**
	Grade 12	0.254	(0.027)**	0.254	(0.027)**	0.240	(0.027)**
	Days Absent	-0.007	(0.000)**	-0.007	(0.000)**	-0.008	(0.000)**
	Low-Performing Student	----	----	0.013	(0.012)	----	----
	High-Performing Student	----	----	-0.009	(0.012)	----	----

Table 4.6 - Biology Results – Continued

		(7)		(8)		(9)	
		Moderator Model - Bio Only - Low Cutoff Tracking - Peer Ability		Moderator Model - Bio Only - Low Cutoff Tracking - Curriculum		Moderator Model - Bio Only - Low Cutoff Tracking - Teacher Characteristics	
Coefficient Group		Coef.	Std. Err	Coef.	Std. Err	Coef.	Std. Err
School-Level	Intercept	-0.276	(0.041)**	-0.318	(0.046)**	-0.347	(0.060)**
	Tracking Quartile - Lowest	0.057	(0.031)	0.057	(0.033)	0.069	(0.032)*
	Tracking Quartile - Medium	0.070	(0.035)*	0.075	(0.036)*	0.087	(0.035)*
	Tracking Quartile - Highest	0.099	(0.037)**	0.105	(0.038)**	0.127	(0.037)**
	Total Per Pupil Exp (excl. Capital & Com. Ser.)	-0.002	(0.000)*	-0.002	(0.001)	-0.002	(0.000)
	Average Daily Membership	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
	Teacher Turnover	0.000	(0.001)	0.001	(0.002)	0.001	(0.001)
	School Pct Black	0.002	(0.001)	0.001	(0.001)	0.001	(0.001)
	School Pct Hispanic	0.005	(0.003)	0.005	(0.003)	0.002	(0.002)
	School Pct Other	0.002	(0.002)	0.003	(0.002)	0.003	(0.002)
	School Pct Free Lunch	-0.001	(0.002)	-0.002	(0.002)	-0.002	(0.002)
	School Pct Reduced Lunch	-0.005	(0.006)	-0.002	(0.006)	-0.006	(0.005)
	Urban Area School	-0.052	(0.042)	-0.055	(0.043)	-0.043	(0.042)
	Rural Area School	0.005	(0.029)	-0.007	(0.030)	0.000	(0.028)

Table 4.6 - Biology Results – Continued

Classroom-Level							
	Peer Dispersion - Read (Std. Dev.)	----	----	----	----	----	----
	Number of Students in Class	-0.002	(0.001)*	-0.001	(0.001)	0.000	(0.001)
	Classroom Percent Black	0.000	(0.000)	-0.001	(0.000)	-0.003	(0.000)**
	Classroom Percent Hispanic	0.000	(0.000)	0.000	(0.000)	-0.002	(0.000)*
	Classroom Percent Other	0.001	(0.000)	0.000	(0.001)	0.001	(0.001)
	Classroom Percent Free Lunch Eligible	-0.001	(0.000)	-0.001	(0.000)**	-0.002	(0.000)**
	Classroom Percent Reduced Lunch Eligible	0.000	(0.000)	0.000	(0.000)	-0.001	(0.000)
	Curriculum Advanced	----	----	0.218	(0.022)**	----	----
	Curriculum Remedial	----	----	0.038	(0.022)	----	----
	Advanced Degree	----	----	----	----	-0.011	(0.018)
	Licensed Other	----	----	----	----	0.026	(0.044)
	National Board Certification	----	----	----	----	0.067	(0.031)*
	Infield Teacher	----	----	----	----	0.091	(0.037)*
	Most or Highly Competitive Barron's Rating	----	----	----	----	0.013	(0.022)
	First Year Teacher	----	----	----	----	-0.144	(0.037)**
	1 to 2 Years Teacher Experience	----	----	----	----	-0.019	(0.026)
	3 to 4 Years Teacher Experience	----	----	----	----	0.017	(0.031)
	10 to 14 Years Teacher Experience	----	----	----	----	-0.024	(0.034)

Table 4.6 - Biology Results – Continued

	15 to 19 Years Teacher Experience	----	----	----	----	-0.044	(0.034)
	20 to 24 Years Teacher Experience	----	----	----	----	0.040	(0.030)
	25 or more Years Teacher Experience	----	----	----	----	-0.065	(0.036)
	Teacher Test Avg (Praxis, NTE, etc.)	----	----	----	----	0.016	(0.015)
	Teacher Non-Certified Pay	----	----	----	----	0.000	(0.000)
Individual-Level							
	Peer Ability Math	----	----	----	----	----	----
	Peer Ability Read	0.166	(0.015)**	----	----	----	----
	Peer Ability Read Squared	0.055	(0.012)**	----	----	----	----
	Peer Ability Read Cubed	----	----	----	----	----	----
	Low-Performing Student * Peer Ability Read	----	----	----	----	----	----
	High-Performing Student * Peer Ability Read	----	----	----	----	----	----
	Std Math Score (Grade 8)	0.323	(0.008)**	0.323	(0.008)**	0.331	(0.008)**
	Std Math Score Squared (Grade 8)	0.041	(0.003)**	0.041	(0.003)**	0.043	(0.003)**
	Std Math Score Cubed (Grade 8)	-0.003	(0.001)	-0.002	(0.001)	-0.003	(0.001)
	Std Read Score (Grade 8)	0.434	(0.007)**	0.429	(0.007)**	0.434	(0.007)**
	Std Read Score Squared (Grade 8)	0.019	(0.003)**	0.018	(0.003)**	0.019	(0.003)**
	Std Read Score Cubed (Grade 8)	-0.020	(0.001)**	-0.020	(0.001)**	-0.021	(0.001)**
	Male	0.121	(0.006)**	0.121	(0.006)**	0.119	(0.006)**

Table 4.6 - Biology Results – Continued

	Black	-0.160	(0.008)**	-0.158	(0.008)**	-0.154	(0.008)**
	Hispanic	-0.020	(0.023)	-0.019	(0.023)	-0.015	(0.023)
	Other	-0.010	(0.014)	-0.008	(0.014)	-0.007	(0.014)
	Underage Student based on Grade	0.143	(0.027)**	0.141	(0.027)**	0.143	(0.027)**
	Overage Student based on Grade	-0.082	(0.008)**	-0.082	(0.008)**	-0.085	(0.008)**
	Academically or Intellectually Gifted	0.071	(0.013)**	0.070	(0.013)**	0.084	(0.013)**
	Disabled Student	-0.067	(0.015)**	-0.074	(0.015)**	-0.070	(0.015)**
	Free Lunch	0.010	(0.008)	0.011	(0.008)	0.014	(0.008)
	Reduced Lunch	0.036	(0.011)**	0.037	(0.011)**	0.039	(0.011)**
	Parent Education Less than High School	-0.014	(0.012)	-0.015	(0.012)	-0.015	(0.012)
	Parent Education Some College	0.063	(0.007)**	0.063	(0.007)**	0.064	(0.007)**
	Parent Education College Graduate	0.075	(0.008)**	0.076	(0.008)**	0.080	(0.008)**
	Parent Education Missing	0.001	(0.073)	0.002	(0.073)	0.002	(0.074)
	Previous LEP Services Recipient	0.012	(0.038)	0.013	(0.038)	0.017	(0.038)
	LEP Services Recipient	-0.069	(0.030)*	-0.070	(0.030)*	-0.069	(0.031)*
	Grade 10	0.057	(0.012)**	0.060	(0.013)**	0.047	(0.012)**
	Grade 11	0.141	(0.018)**	0.143	(0.018)**	0.126	(0.018)**
	Grade 12	0.251	(0.027)**	0.254	(0.027)**	0.238	(0.027)**
	Days Absent	-0.007	(0.000)**	-0.007	(0.000)**	-0.008	(0.000)**
	Low-Performing Student	----	----	----	----	----	----

Table 4.6 - Biology Results – Continued

	High-Performing Student	----	----	----	----	----	----

Peer Ability and Biology Achievement

The series of models utilized to investigate the stated hypotheses in Biology use the test performance of peers on 8th grade reading tests to indicate peer ability. Research Question 1 predicts that students with higher performing peers (measured by averaging the 8th grade reading test performance of all of the other students in a classroom) will score better on the Biology exam compared to similar students in classrooms with lower performing peers. Research Question 1 is supported in the model results 'Biology with Peer Effects – Reading Only' (Table 4.6 Model (2)). Placing a student in a class where one's peers scored one standard deviation higher than average is predicted to increase that student's performance on the EOC Biology exam by about 0.09 standard deviations (Table 4.6 Model (2)). Similarly, students placed in classrooms with lower performing peers are expected to perform worse. Model (3) tests for any non-linear impacts of peer ability on a student's Biology test performance (Table 4.6). Based on this model, the influence of peer ability on Biology test performance increases at an increasing rate over the range of relevant peer ability scores. For a student with peers whose scores were 2 standard deviations above average, the predicted impact on the student's Biology score is an increase of about 0.26 standard deviations compared to a similar student in a classroom whose peers performed at the mean in reading in the prior year (Table 4.6 Model (3)). The third order term on prior peer achievement was not statically significant in Model (3) and is excluded from subsequent models (Table 4.6).

Peer Dispersion and Biology Achievement

Research Question 2 predicts that students in more academically diverse classrooms (those where the dispersion of prior reading test performance is greater) will score better on Biology exams compared to similar students in less academically diverse classrooms. The results of Model (4) lead to a rejection of RQ 2 as there is no statistically significant difference in test performance between similar students in classrooms that are more or less academically dispersed in Biology classrooms (Table 4.6).

The final model, focused on peer differences, tests whether the impact of peer ability on student test performance differs for students in the top or bottom fifth of the distribution in either reading or math based on 8th grade test performance. In Model (5), I find that lower performing students appear to differentially benefit from placement in a classroom with higher performing peers (Table 4.6). A low-performing student (defined as a student scoring in the bottom 20% of the distribution of scores in 8th grade reading or math) is expected to gain an additional 0.07 standard deviation units when placed in a class with peers who scored at the mean on 8th grade reading exams (Table 4.6 Model (5)). This predicted gain is in addition to the regular gain for higher performing peers (about 0.08 + 0.05 or 0.13). While the returns to students from peer ability might differ, the student must have scored lower in 8th grade to be considered a 'low performing' student. The returns from peer ability are similar from students in the middle three fifths of the distribution in 8th grade score and students in the top twenty percent.

Tracking and Biology Achievement

For students enrolled in Biology, I find no differences between students in schools with no evidence of tracking and schools with the lowest levels of tracking intensity. In support of Research Question 3, I do find positive and statistically significant differences between students in schools with high or medium tracking intensity and comparable students in schools with low or no tracking (Table 4.6 Model (6)). Students in schools with medium tracking intensity scored on average about 0.09 standard deviations higher than students in schools with lower levels of tracking or no tracking holding other characteristics constant (Table 4.6 Model (6)). In schools with the high tracking intensity, students scored about 0.12 standard deviations higher than similar students in schools with low to no tracking (Table 4.6 Model (6)). Models (7) to (9) test three possible sets of variables expected to mediate the tracking intensity effects, peer ability, curriculum differences, and teacher characteristics. For Biology students, both peer ability and curriculum differences are found to be partial mediators of tracking effects. In the peer ability mediator, Model (7), both medium and high tracking intensity schools' coefficients are reduced when peer ability is added to the model (Table 4.6). For medium tracking intensity schools, the coefficient is reduced from 0.088 to 0.07, a reduction of about 19 percent (Table 4.6 Model (7)). For the most highly tracked schools, the coefficient on tracking intensity is reduced from 0.012 to about 0.1, a decrease of about 19 percent as well (Table 4.6 Model (7)). In Model (8), testing mediator effects from curriculum differences also reduces the magnitude of coefficients on students enrolled in schools with medium and high tracking intensity (Table 4.6). The reductions in this case are about 15 percent for both coefficients (Table 4.6 Model (8)). As in the model testing the mediating effects of tracking intensity on Algebra I students, no evidence suggests that teacher characteristics mediate the impact of tracking on Biology students' EOC test scores (Table 4.6 Model (9)). In Model (9), the coefficient on low intensity tracked schools becomes statistically significant with a coefficient of 0.069. In addition to the model results presented here, one additional model integrating all three sets of possible mediators in Biology failed to mediate the effects of tracking intensity any further than the individual mediator models shown.

Biology Control Variables

Model (1) for Biology students is consistent with prior results in other subjects where individual student characteristics are more predictive of student outcomes compared to characteristics at the school or classroom level (Table 4.6). Student performance on 8^{th} grade end-of-grade exams in reading are the strongest predictor of Biology test performance. In the Model (1), end-of-grade tests from grade 8 in both reading and math had statistically significant linear and non-linear effects (Table 4.6).

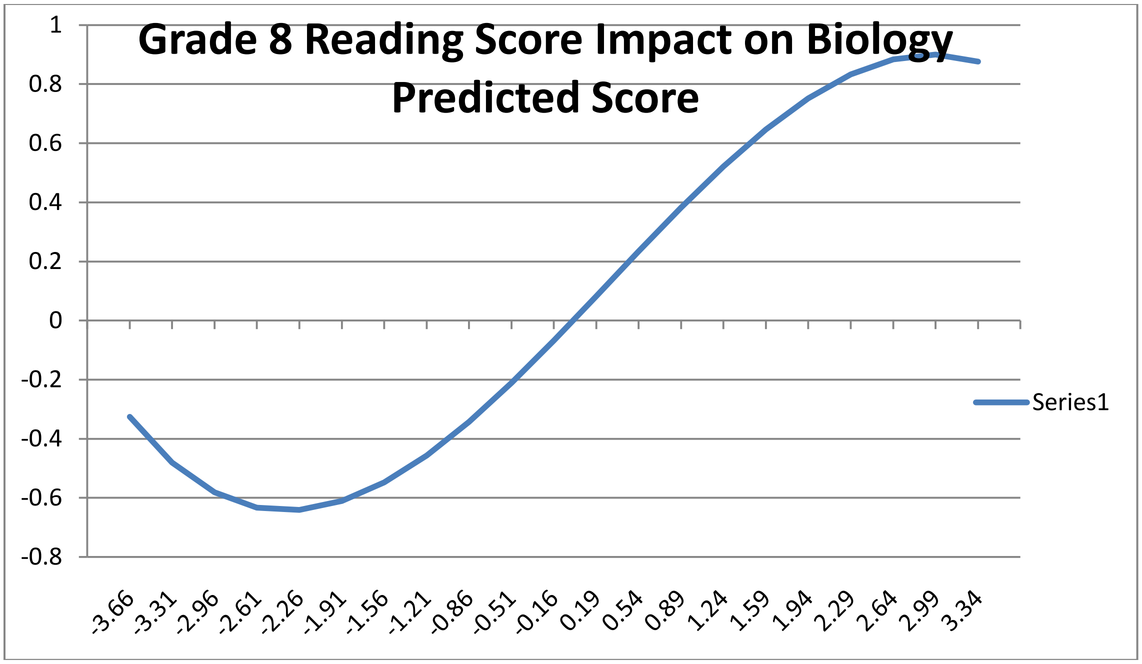

Figure 4.5 – Grade 8 Reading Score Impact on Biology Predicted Score

Figure 4.5 provides a graphical representation of the impact of 8^{th} grade reading test performance over the relevant range of test scores on the Biology end-of-course test. The graph shows that once the prior reading performance reaches about 3 standard deviations above the mean, the maximum positive effect is reached (about 0.90 standard deviation units). For 8^{th} grade scores which were below average, lower test performance is predicted, but the maximum negative impact is estimated to be where the prior reading test score is about two and one quarter standard deviations below zero (-2.24). Students scoring at this level on the 8^{th} grade reading exam are expected to score about 0.64 standard deviations below comparable peers on the Biology end-of-course exam. Statistical significance on these three terms, Std Read Score (Grade 8), Std Read ScoreSquared (Grade 8), and Std Read Score Cubed (Grade 8), is maintained across models and they remain relatively stable across models.

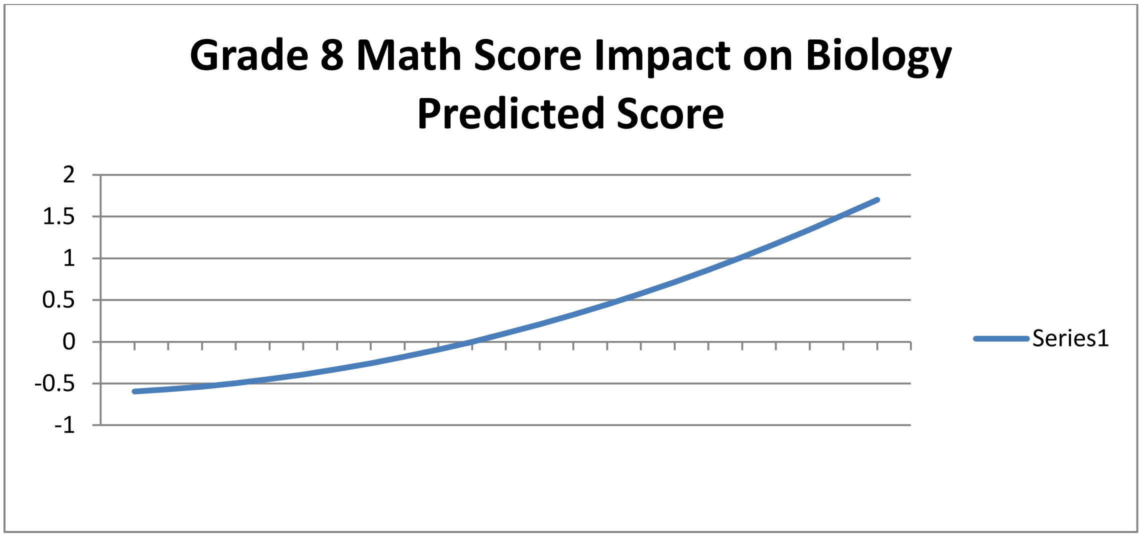

Figure 4.6 – Grade 8 Math Score Impact on Biology Predicted Score

Figure 4.6 shows the expected impact of prior performance on 8th grade math scores on Biology test performance. Only the Std Math Score (Grade 8) and Std Math Score Squared (Grade 8) were statistically significant across all of the peer effects related models in Biology. As prior math performance increases, the influence on Biology test performance increases at a slightly increasing rate. At the highest value of student prescores, about 3.6 standard deviations above zero, the student is expected to score about 1.7 standard deviations higher on the Biology exam compared to a similar student who scored at the mean (0) on their 8th grade math exam. At the bottom of the distribution of prior math scores, the effect flattens with students with the lowest scores (3 standard deviations below zero) predicted to score only about 0.6 standard deviations lower on the Biology end-of-course exam. Calculations of score impacts are based on Model (1) results (Table 4.6).

On average, male students perform better than similar female students, scoring about 0.12 standard deviations higher than comparable female students (Table 4.6 Model

(1)). White, Hispanic, and students of other ethnicity perform similarly on the Biology exam, while Black students substantially worse (0.16 standard deviations lower) (Table 4.6 Model (1)). Underage students score higher (0.14 standard deviations) and overage students perform lower (about 0.08 standard deviations) than similar students who are aged within the cutoffs for their grade (Table 4.6 Model (1)). Students coded as academically or intellectually gifted perform, on average, better than similar students (about 0.07 standard deviations), while those coded as disabled perform worse than non-exceptionally coded students (about 0.07 standard deviations) (Table 4.6 Model (1)).

Biology test takers coded as receiving reduced lunches score slightly higher, on average, than those paying full price for lunch (about 0.38 standard deviations) (Table 4.6 Model (1)). In Biology models, there is no statistically significant difference in performance between free lunch students and those coded as full pay (Table 4.6). Parental education also predicts test performance at statistically significant levels for some groups of students. Compared to students with parents who are high school graduates, students with parents who attended some college or are college graduates performed better than comparable students (0.06 and 0.07 standard deviations higher respectively) (Table 4.6 Model (1)). Students whose parent educational level was coded as missing or did not finish high school did not significantly differ from those whose parents were high school graduates (Table 4.6).

Students receiving services as Limited English Proficiency students scored lower than comparable students who never received these services (about 0.07 standard deviations lower) (Table 4.6 Model (1)). There were not statistically significant differences between students coded as former LEP services recipients and other students.

Compared to 9th grade Biology students, students in higher grades perform better on the Biology EOC exam with students in the highest grades performing best. On average and compared to 9th grade students, 12th grade students score about 0.25 standard deviations higher, 11th grade students score about 0.14 standard deviations higher, and 10th grade students score about 0.06 standard deviations higher compared to similar students (Table 4.6 Model (1)). As expected, missing school days leads to decreased performance on the EOC Biology exam, each additional day of school missed decreases a student's exam score by about 0.007 standard deviation units compared to a similar student with one fewer days absent. A student missing 10 days would be expected to score about 0.07 standard deviations lower than a similar student who did not miss any school days (Table 4.6 Model (1)).

The base model's classroom characteristic variables reveal a few differences in student outcomes as a result of different classroom conditions and the teacher's characteristics within them. While Model (1) predicts no differences in achievement based on class size, subsequent peer effects models predict slightly worse performance in larger classrooms (Table 4.6). Decreasing a class size by one student below the average class size, is expected to increase student outcomes by about 0.003 standard deviations compared to students in similar classrooms (Table 4.6 Models (2) – (5)).

None of the observed peer effects models predict differential performance based on ethnic composition. In Model (1), slightly lower student performance is predicted in classrooms with larger concentrations of students eligible for the free lunch program (Table 4.6). The impact, however, is quite small, about 0.001 standard deviation units, for a one percent increase in a classroom's proportion of free lunch students compared to the mean classroom (Table 4.6 Model (1)). The mean classroom consisted of about 25% free lunch eligible students.

Students in Biology classrooms coded as providing an advanced curriculum perform significantly better in the base model (0.22 standard deviations higher on average) than similar students in classrooms providing the regular curriculum (neither advanced nor remedial) (Table 4.6 Model (1)). The impact of curriculum is diminished once the ability level of a student's peers is taken into account. Students enrolled in remedial curriculum classrooms perform no differently on average than those in regular instruction classrooms in Model (1), but subsequent models predict higher achievement in remedial level classrooms compared to the reference group (0.04 standard deviation units) (Table 4.6).

Teacher characteristics impacting student outcomes in Model (1) of Biology performance include the teacher's infield status and years of experience (Table 4.6). Students of infield teachers, those with continuing or initial licenses in High School Biology or Science, performed better (about 1/10 of a standard deviation in the base model) than similar students in classrooms taught by an instructor who was not certified to teach High School Biology with an initial or continuing license (Table 4.6 Model (1)). Students in classrooms whose teachers were in their first year of teaching or with teachers having more than 25 years of experience performed worse (about 0.14 and 0.08 standard deviations) on average, compared similar students whose teachers had levels of experience between these two extremes (Table 4.6 Model (1)). These statistically significant teacher characteristics remain stable and significant across the peer effects models in Biology.

These models also include school characteristics that might influence the Biology test performance of students. Comparing similar students and classrooms in different school settings, students within schools that spend more than average score slightly worse than similar students in average spending schools. A student in a school whose total per pupil expenditures are $1,000 dollars higher than average is expected to score about 0.002 standard deviations lower in Biology than a similar student (Table 4.6 Model (1)). No other school-level characteristics were statistically significant predictors of student performance on the Biology EOC exam.

Comparisons Across Subjects

Table 4.7 – All Subjects Summary Comparison

Research Question Area	English I	Algebra I	Biology
RQ1: Peer Ability	Positive	Positive	Positive
Non-Linear Peer Influence	Increasing Returns	No Effect	Increasing Returns
High Performing / Low Performing Student	No Effect	No Effect	No Effect
High Performing * Peer Ability Interaction	No Effect	Positive	No Effect
Low Performing * Peer Ability Interaction	No Effect	No Effect	Positive
RQ2: Peer Dispersion	No Effect	No Effect	No Effect
RQ3: Tracking - Lowest Quartile	No Effect	No Effect	No Effect
RQ3: Tracking - Medium Quartile	No Effect	No Effect	Positive
RQ3: Tracking - Highest Quartile	No Effect	Strong Positive	Strong Positive
Tracking Mediators	N/A	Peer Ability	Peer Ability & Advanced Curriculum

Table 4.7 summarizes the previously reported coefficients across subjects to more easily compare the results. I replace the numeric values previously reported with descriptive indicators of the influence of the variables of interest on student EOC scores in each subject. Research Question 1 suggests a positive relationship between the ability level of peers and student EOC test score outcomes. In all three subjects, there are strong associations between the ability level of peers, as measured by prior performance on 8th grade end-of-course exams in a related subject, and individual high school end-of-course exams (Table 4.4). These analyses support Research Question 1 across all three subjects. In addition, I find support for non-linear peer ability influences for students enrolled in English I and Biology. Over the relevant ranges of peer ability values, the relationship is increasing at an increasing rate. While the sample size for the Algebra I model is smaller, this does not appear to be an explanation for a lack of relationship between squared and cubed versions of the peer ability variable as the standard errors are approximately equal across subjects. The analysis for Algebra I differs slightly due to the compressed nature of prior math achievement since a number of higher achieving 7th graders enroll in Algebra I during the 8th grade. The use of the inverse mills ratio weighting in the Algebra I models addresses more significant selection threats to the validity of the study, but the different enrollment pattern for Algebra I high school students causes decreased variance in the independent variable measuring peer ability. This decreased variance may limit the ability of the analysis to detect any non-linear relationship between peer ability and individual student EOC test score outcomes.

Research Question 2 examines the relationship between student end-of-course exam outcomes and the range of student abilities within a classroom. In all three subjects, Research Question 2 is rejected as all analyses found no relationship between these variables different from zero. The final peer effects related models looked closely at the relationship between peer ability and students in the top or bottom twenty percent of test scorers based on 8^{th} grade end-of-grade tests in reading or math. Students scoring in the top twenty percent of test takers in reading or math were coded as High Performing, while those in the bottom twenty percent in reading or math were coded as Low Performing. For English I students, no additional statistically significant relationships were revealed. In Algebra I, gains were higher on average for high performing students when placed in classrooms with higher ability peers compared to similar students who were not in the top twenty percent of test performers in 8^{th} grade. In Biology, the analysis revealed additional positive gains for low performing students when placed in classes with higher performing students compared to similar students that scored higher on 8^{th} grade exams. Overall, I find two statistically significant relationships between the interactions of peer ability and a student's high or low performing status. High performing students are associated with additional gains when placed in classrooms with higher performing peers, while low performing students are associated with additional gains when placed in classrooms with higher performing peers. In four other cases, no other statistically significant coefficients on these interaction variables were found.

Research Question 3 focused on the relationship between school tracking intensity and student test score outcomes for end-of-course tests in three subjects. The model results suggest that English I test outcomes are not improved in schools with various levels of tracking intensity. In addition, no statistically significant relationship is revealed between low levels of tracking and student test performance in Algebra I or Biology

(Table 4.4). Some positive relationship between medium tracking intensity and Biology test performance exists, but no relationship is observed in Algebra I or English I medium intensity tracked schools. In both Algebra I and Biology, I observe a positive and statistically significant relationship between student end-of-course test score outcomes in high intensity tracking schools (Table 4.4). In further models to test mediators of tracking, I find that peer ability is a consistent mediator of tracking, exhibits mixed results for curriculum, and shows no mediating relationship between high intensity tracked schools and teacher characteristics. The mediator variables explained only about 20 percent of the positive relationship between high intensity tracked schools and student test score outcomes. In summary, I find that only the most highly tracked schools appear to generate substantial positive associations with student test score performance in high school and these positive effects are limited to only two of the three subjects studied in this analysis.

CHAPTER 5

DISCUSSION

Introduction

The purpose of this dissertation is to add to the body of evidence regarding the assignment of students to classrooms at the high school level. The dissertation informs our thinking about the impact of prior peer test performance on individual test score achievement in three different high school subjects (English I, Algebra I, and Biology). In addition, the findings provide new information on the relationship between school tracking and student achievement within high schools.

While peer effects and tracking have been an interest of researchers in a variety of fields over recent decades, the opportunity to examine this issue at the high school level with a rich individual-level dataset is unique. Further contributions are the analysis of student performance in three high school subjects including a science course, Biology, going beyond the usual focus on mathematics and reading test scores and creation of a tracking measure that treats each subject separately based on observed prior test score distributions. Finally, the analysis utilizes hierarchical linear modeling to appropriately calculate standard errors to account for clustering of students within classrooms within schools.

Interpreting the Effects of Peers: How Substantial are the Effects?

English I models estimated the relationship between peer ability and EOC test score outcomes as coefficients of 0.124 standard deviation units on the peer ability variable and 0.016 standard deviation units on the peer ability squared variable. When estimating a total effect of this coefficient for a student who is placed with peers scoring one standard deviation above average on their 8th grade reading exams, one sums the two coefficients for a total of 0.14 standard deviations (0.124 + 0.016). While the observed coefficients on peer effects appear large, the actual predicted impact on overall student test score outcomes is modest for the majority of students since it is not possible to place all students with peers who scored far above or below the mean.

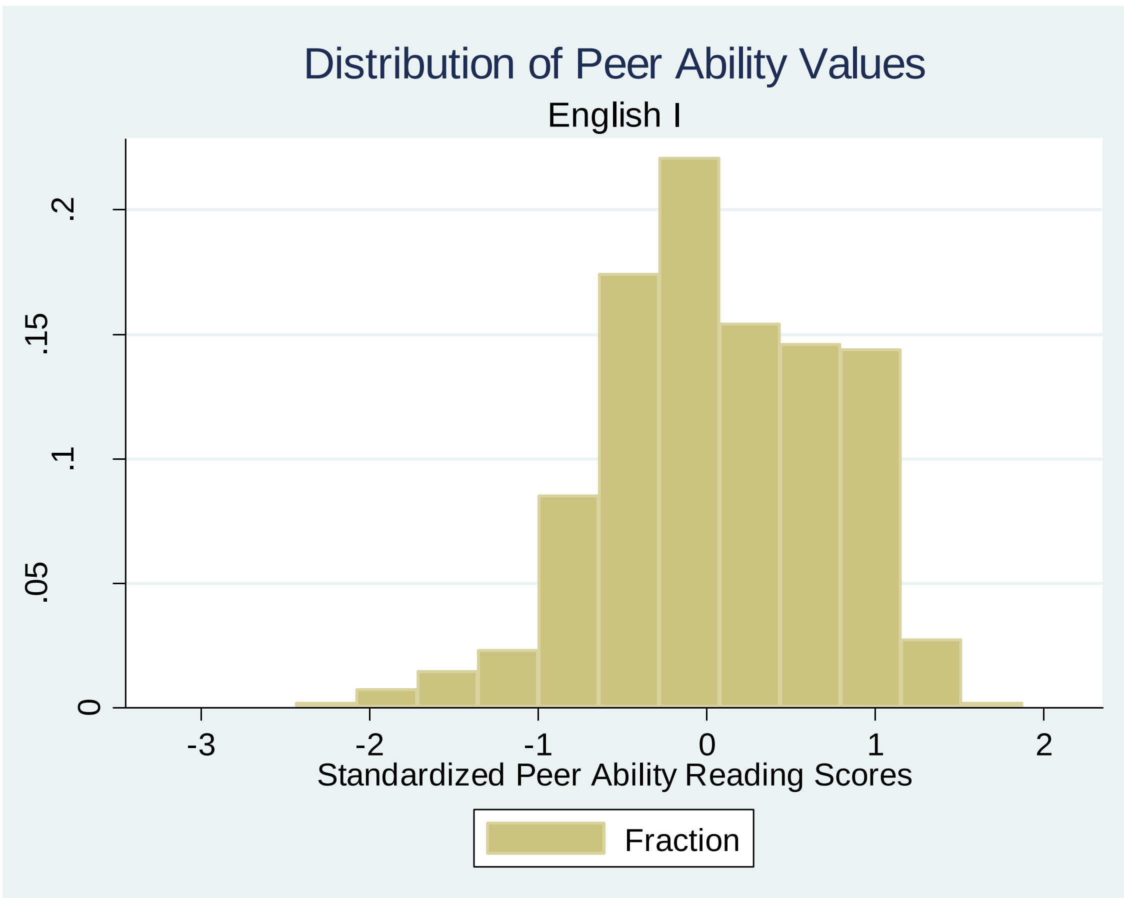

Figure 5.1 – Distribution of Peer Ability Values – English I

Figure 5.1 above displays the proportion of students with peer values in the specified range for English I students. As expected, average values for peer ability cluster around zero. This distribution means that for most students the estimated impact of peer ability on their EOC test score outcome will be near zero for coefficients that are less than 0.5. Standardized peer ability reading values appear on the x-axis and the proportion of student scores in the displayed range appear on the y-axis.

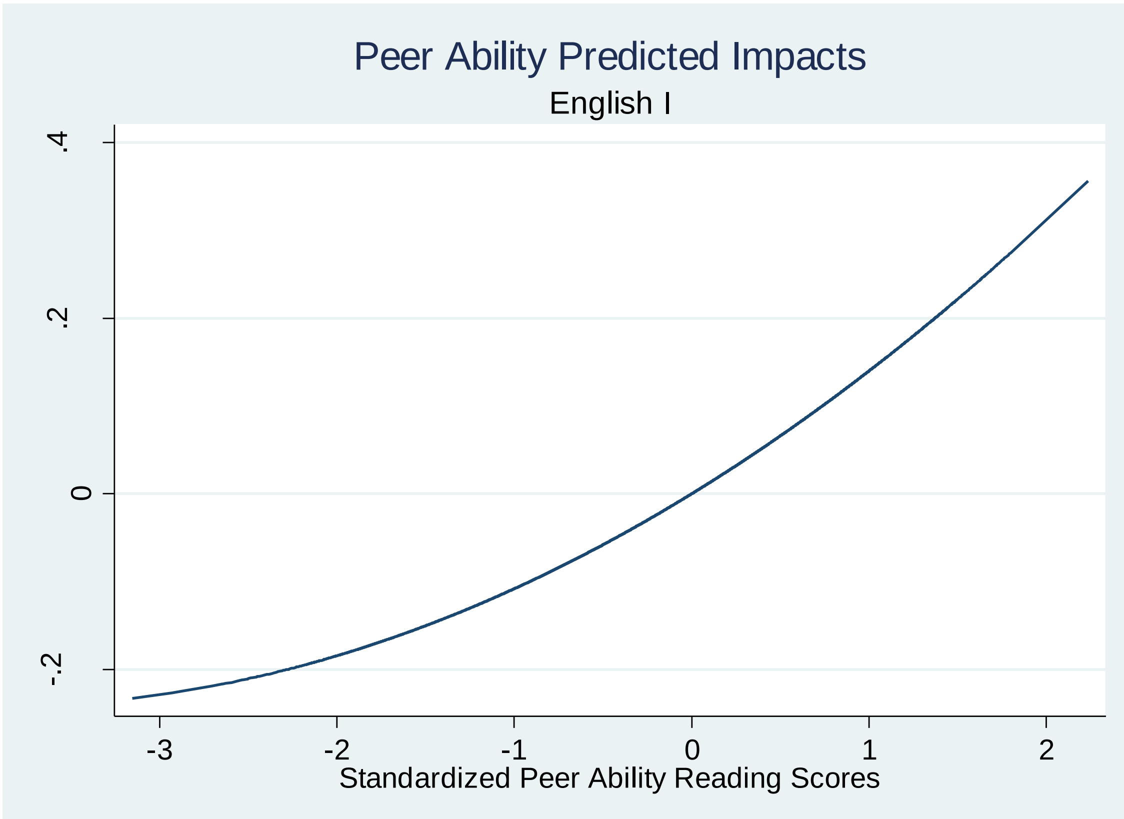

Figure 5.2 – Peer Ability Predicted Impacts – English I

Figure 5.2 plots the relationship between peer ability and student outcomes on the English I End-of-Course exam. For example, a student with peers that score 1 standard deviation below zero is expected to score about -0.1237 (0.124 * -1 + 0.016 * $(-1)^2$) standard deviations lower compared to a similar student with peers scoring at the mean (zero) on end-of-grade (EOG) reading tests from 8^{th} grade. A student with peers that score 1.5 standard deviations above average is expected to score 0.222 (0.124 * 1.5 + 0.016 * $(1.5)^2$) standard deviations higher compared to a similar student with average performing peers. The x-axis remains the same as Figure 5.1, the standardized peer ability value for a student, but the y-axis now represents the estimated change in English I

EOC score based on coefficients in English I non-linear peer effects models presented in Chapter 4. The line displayed represents the predicted change in EOC score for a student with peers of a given value.

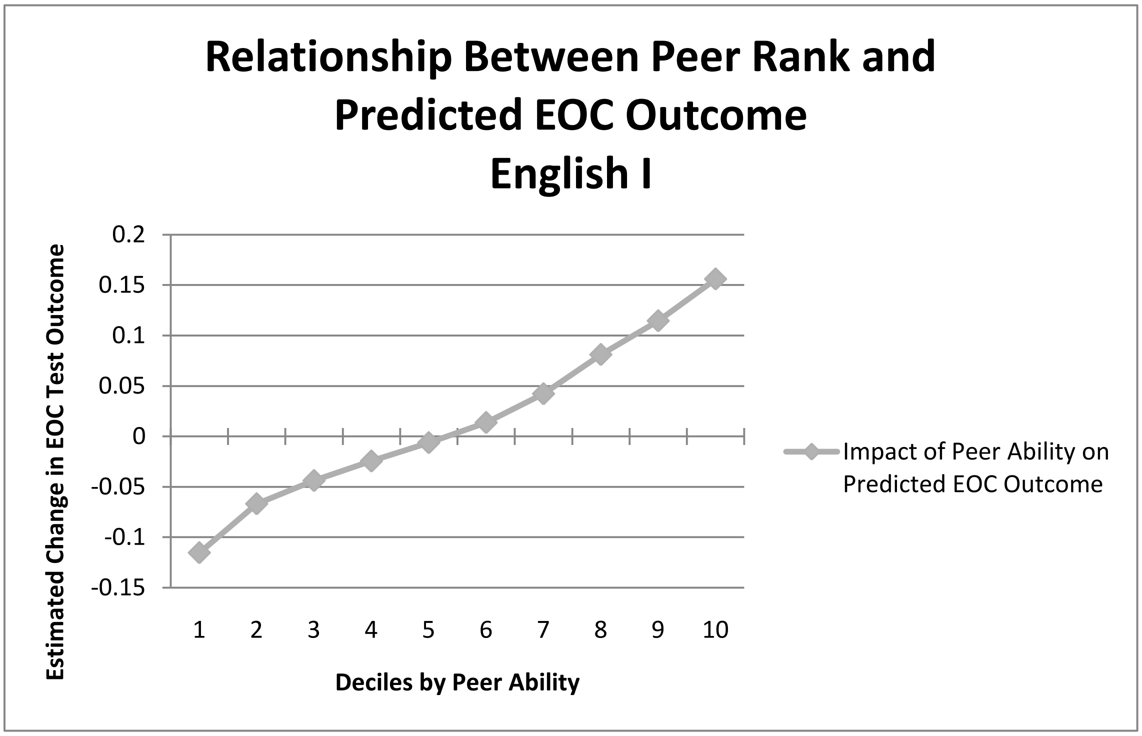

Figure 5.3 – Peer Rank and Predicted EOC Outcome – English I

Figure 5.3 plots the estimated change in EOC test score outcomes for each decile of peer ability. Each decile represents ten percent of the students in the data set based on their values of peer ability. The figure shows that for students in the 3rd through 7th deciles, the predicted effect of peer ability is between -0.05 and 0.05. These differences compare to an effect size of 0.082 standard deviations from being enrolled in an advanced curriculum English I class or a negative effect of -.05 standard deviations from missing ten days of school. For the average student, the estimated impact of peer ability on English I test score outcomes is 0.015 standard deviations (the mean reading peer ability * the coefficient on peer ability + mean reading peer ability squared * the coefficient on peer ability squared). Students in the first and 10th deciles of students based on prior peer reading performance are expected to score about 0.11 standard deviations lower and 0.156 standard deviations higher respectively on English I EOC exams.

For the average student, the impact of peer ability on student test score outcomes in high school Algebra I is about -0.029 standard deviations. The peer effect for Algebra I high school students is linear in nature based on the analysis results in Chapter 4.

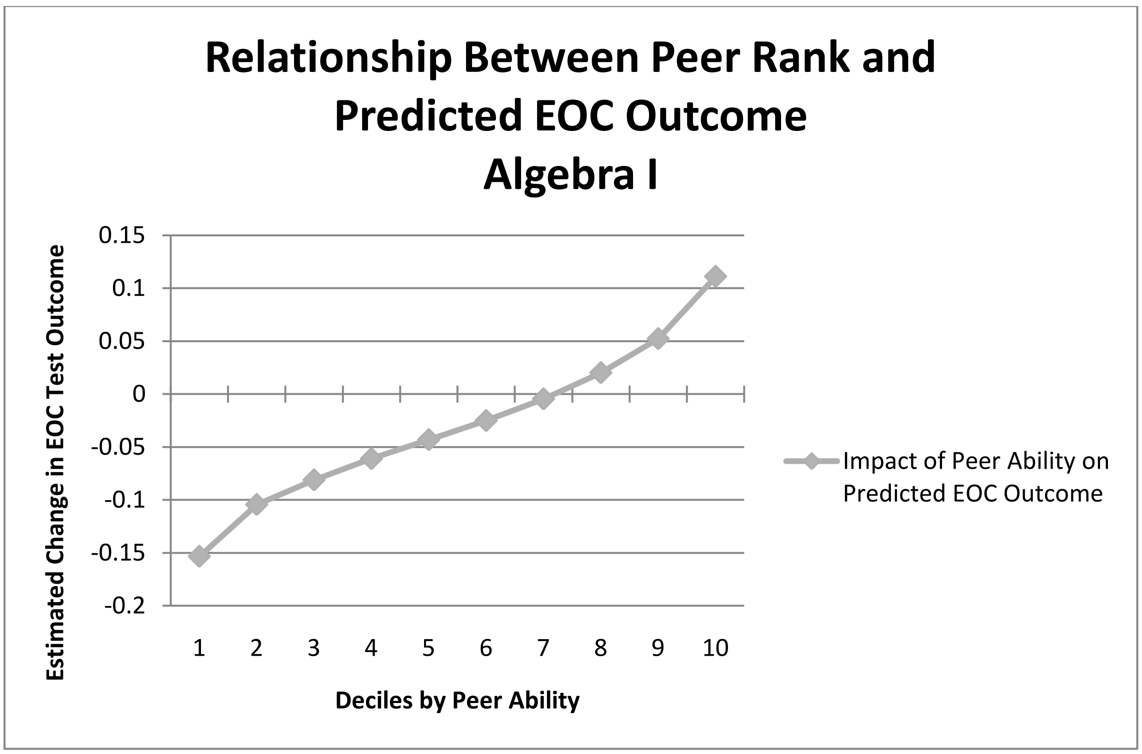

Figure 5.4 – Peer Rank and Predicted EOC Outcome – Algebra I

Similar to the results in English I, for most students, the impact of peers on test score outcomes is small. Students whose peers are in the 4th through 9th deciles based on prior 8th grade math achievement range from a low estimated effect of -.061 standard deviations to a high of 0.052 for student in the 9th decile. Only students in the 1st, 2nd, and 10th decile have estimated effects that are more than 1/10 of a standard deviation away from zero impact. Students placed in classrooms with the lowest performing peers, the 1st decile, are expected to score about -0.153 standard deviations lower than comparable students while those placed with the highest performing peers are expected to score about 0.111 standard deviations higher than their comparable peers. These effects at the extremes of the distribution are larger than those observed for students taught Algebra I by Nationally Board Certified teachers (0.083 standard deviations higher) or those taught by first year teachers (0.078 standard deviations lower). Models run on high performing students (those who scored in the top 20 percent of the test score distribution on their math exams in 8th grade) showed an additional gain of 0.06 standard deviations on expected Algebra I EOC test performance.

Note that inclusion of the inverse mills ratios to weight cases for selection into 8th grade Algebra I classes does not adjust the distribution of peer ability levels for students enrolled in these classes. Because the highest performing students in 7th grade were enrolled in Algebra I during their 8th grade year, high school Algebra I students are more likely to be enrolled in classrooms with lower performing peers based on 8th grade EOG math exams. Figure 5.4 shows this distributional effect as only the top three deciles of student peer ability values were above zero. If all students took Algebra I for the first time in a single grade, the ability levels of peers would be more evenly distributed reflecting a distribution that included more higher performing peers. The slight negative effect of overall peer ability on student outcomes for these students may be offset by a small positive peer ability effect for students enrolled in Algebra I in 8th grade. This explanation is consistent with the findings but was not tested directly.

The expected distribution of peer ability impacts for Biology students mimics the results from the English I analysis where the line flattens as peer values get farther away from zero (Figure 5.4).

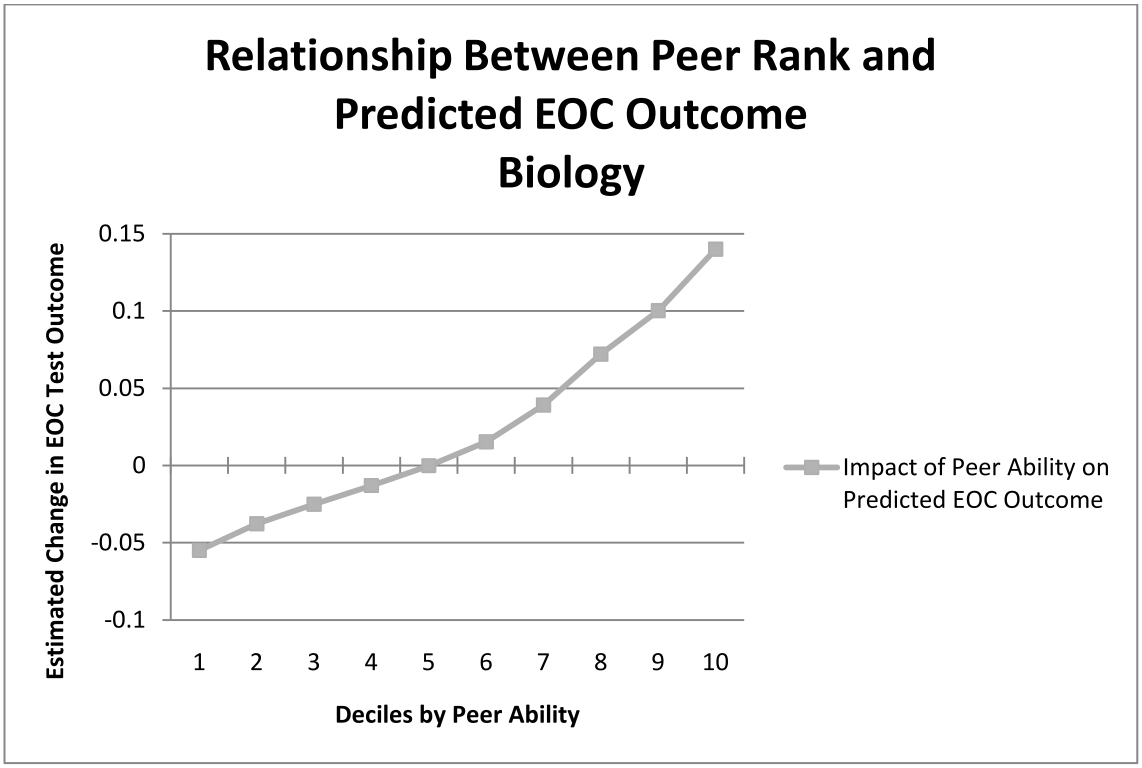

Figure 5.5 – Peer Rank and Predicted EOC Outcome – Biology

Students within the lowest decile of reading peers score about 0.055 standard deviations lower than similar students with peers scoring at the mean on end-of-grade 8 exams in reading. For students in the 2nd through 7th deciles, the association between Biology EOC outcomes and peer reading ability is between -0.05 and 0.05. Students with the highest scoring peers, the 10th decile, are associated with gains of about 0.140 standard deviations. By comparison, Biology students taught by infield teachers compared to those with teachers not holding initial or continuing licensure to teach Biology or High School Science score about 0.08 standard deviations higher than comparable peers, and students with first year teachers in Biology classrooms score about 0.078 standard deviations lower than comparable students taught by teachers with five to nine years of experience.

In addition to these general findings on peers, the study also showed that lower performing Biology students (those scoring in the bottom 20 percent of 8th grade students in reading or math) experienced differential peer effects from other Biology students. Low performing students experienced an additional 0.074 standard deviation estimated gain based on peer ability. A graphical version of this would shift the solid line in Figure 5.5 up by 0.074 standard deviations corresponding to the expected peer reading ability impact at each point in the distribution for low performing students placed in classrooms with peers in the specified decile.

While each subject was unique in its relationship between prior peer ability and student EOC test score outcomes, the general pattern of a positive and statistically significant relationship was consistent. In addition, the association between prior peer ability and student EOC test score outcomes is positive for students with peers whose eighth grade test performance was above average and negative for students with peers whose eighth grade test performance was below average. Finally, the relationships were found to be non-linear for both English I and Biology peers. In both of these subjects, the predicted impact of peer ability increases at an increasing rate for peer ability values greater than -1.00 standard deviations.

In addition to estimating the impact of peer ability on individual student outcomes, a second goal of the analysis was to investigate potential links between the dispersion of peer abilities within a classroom and student outcomes on EOC exams.

Across all three subjects, no models demonstrated a statistically significant relationship between the standard deviation of grade 8 standardized scores within a classroom and end-of-course exam outcomes. This result indicates that for high school students in the three subjects examined, there is no positive or negative association with EOC test score outcomes for classroom assignments that attempt to maximize student heterogeneity in terms of ability. This finding contrasts with the Vigdor and Nechyba (2004) and Zimmer and Toma (2000) findings of positive relationships between peer dispersion and student test score outcomes. This may be due to differences in the maturation level of students in the prior studies compared to this one (the previous analyses utilized data on younger students), possible bias related to only a single measure of prior achievement in the Zimmer and Toma study, or omitted variable bias which overstated the relationship between peer dispersion and student test score outcomes. An additional difference included the inclusion of roster entries to conduct teacher-student matches in this study compared to the test administrator utilized in Vigdor and Nechyba (2004).

Finally, the analysis sought to quantify the relationship between school tracking intensity and student outcomes on end-of-course exams. No relationship between the two was observed with English I EOC scores as the outcome variable. In models where Algebra I EOC scores were the outcome variable, schools with the highest levels of tracking were associated with scores about 0.147 standard deviations higher than comparable students in schools with less or no tracking. Similarly, Biology students in schools engaging in moderate or high levels of tracking intensity were also associated with higher EOC test score performance, about 0.088 standard deviations and 0.124 standard deviations respectively. While these positive effects in Biology and Algebra I were partially explained by including controls for peer ability level in models, this relationship did not fully mediate the positive effects of tracking.

This Study in the Research Context

Table 5.1 – Updated Peer Achievement Literature [7]

	Grades; Location	Estimated Impacts - Standardized Estimate of Peer Achievement
Zimmer and Toma 2000	Ages 13-14 (8th); International including USA	0.15
Hanushek, Kain, Markman, and Rivkin 2003	3 – 6; Texas	0.27 to 0.43
Vigdor and Nechyba 2004	5; North Carolina	0.05 to 0.08
Hoxby and Weingarth 2006	3 – 8; Wake County, North Carolina	0.25 (linear-in-means model)
Burke and Sass 2008	3 – 10; Florida	0.04 (HS Reading) 0.06 (HS Math) (with Teacher FE)
Zabel 2008	3 – 4; New York City	0.04 to 0.08 (student fixed effects models)
Fortner 2010	9 – 12; North Carolina	0.09 to 0.13 (Algebra I, English I, and Biology)

Table 5.1 above places this study into a table of selected prior literature focused on peer effects. While numerous studies have tackled these issues of peer effects and tracking, they have largely focused on one or the other of these topics and on math and reading performance of students in elementary or middle grades. One contribution of this dissertation is the investigation of peer effects at the high school level. In addition, this study is the first to include information on peer effects in specific high school subjects including a science course (Biology). I also develop a novel method for classifying schools by tracking intensity and use this classification scheme to test the extent to which prior peer achievement mediates the observed positive association between tracking

[7] Adapted from Table 2 in Zabel (2008, 208-9).

intensity and student test score achievement. Testing observed teacher characteristics and curriculum level as mediators of tracking intensity are further contributions of this study.

This study generated linear peer effects estimates of about 0.09 to 0.13 standard deviation units depending on the subject. These estimates lie roughly between the high estimates of Hanushek, et. al. (2003) and Hoxby and Weingarth (2006) and the low estimates of Vigdor and Nechyba (2004), Burke and Sass (2008), and Zabel (2008). Estimates from Burke and Sass (2008) and Zabel (2008) included classroom level fixed effects (FE) estimates which may have contributed to the low peer effects estimates observed in those studies. Burke and Sass (2008) run models both with and without teacher FE and find generally that peer effects coefficients are higher when teacher FE are omitted, but for high school math, coefficients are actually reduced when teacher FE are included in models (38). The field would benefit from further work in this area which adds additional validity to the estimates of high school student peer ability effects, and efforts to determine to what extent the differences observed in estimated effects are due to model specification, more extensive control variables, or merely differences related to subject and grade level.

The results of models testing the association between peer ability dispersion and student test score outcomes implies that efforts to alter student test score performance via this mechanism will not generate results for high school students in the subjects examined as a part of this study. None of the models generated statistically significant associations between peer ability dispersion and student test scores. With widely differing estimates of the relationship between peer ability heterogeneity and student outcomes, additional studies can contribute through efforts to stabilize estimates of this relationship or recognition of real differences based on grade level and/or subject (Zimmer and Toma, 2002; Hanushek, et al., 2003; Vigdor and Nechyba, 2004; Zabel, 2008). Continual improvements in research designs can also inform the differences across subjects and grades.

Among the hypotheses pursued in this dissertation, the tracking related section provides the most intriguing area for further study. Prior evidence supports hypotheses that tracking or ability grouping has no general effect on student achievement (Betts and Shkolnik (2000a), is helpful to low performing students (Figlio and Page, 2002), and is harmful to low performing students (Argys et al., 1996). These contradictory studies examine the issue from different perspectives, but all use the reports of school administrators and teachers to determine the definition of tracking. In this dissertation, I use a definition of ability tracking that is based on the data itself and find that high intensity tracking is associated with higher student EOC test performance in Algebra I and Biology courses. Also, medium intensity tracking is associated with higher student EOC test performance in Biology only. Low intensity tracking and all ranges of tracking intensity in English I produced no statistically significant effect on student EOC test score outcomes.

Additional efforts to determine mechanisms of tracking would be very useful in determining what behaviors or organizational characteristics are likely to contribute to the observed relationships between tracking intensity and student test score outcomes. Additional hypotheses related to the impact of tracking on high and low performing students would be of immediate benefit to educational policymakers informing aspects of the equity and efficiency trade off as they relate to tracking intensity. There is also an opportunity for scholars to seek to understand differences in the way that English I and Algebra I and Biology classroom instruction differs that might account for the divergent results (other than purely subject differences). For example, individual drill and practice in Algebra I may contrast with group oriented lab work in Biology or classroom discussion methods of instruction in English I. Also, I plan to pursue publication of tracking analyses with the novel approach to defining tracking which avoids the conflict between suggestions of informal and formal tracking encountered in the critiques of previous work by Argys, et al. (1996) and Betts and Shkolnik (2000a). Further research should incorporate recent critiques of value added used in this and other studies to confirm that the results are similar when approached with additional prior test performance results or student fixed effects (Rothstein, 2009). Fixed effects approaches may be complicated when EOC test scores serve as outcomes. The three subjects included in this study are largely universally taken by high school students, but other science and math courses are complicated by selection. Only particular students will take higher level math and science course like Geometry, Physics, or Chemistry. These students may differ substantially from students who take only the minimum EOC courses required for graduation. While experimental evidence of effects is often preferred, the controversial nature of ability tracking may be an area where quasi-experimental or evidence from ‘natural’ experiments may provide better information as the subjects of tracking experiments may significantly alter their behavior when they know that they are subjects of study.

Policy Implications

While this analysis does confirm that the overall impact of peers is relatively small when viewed on an aggregate basis, the estimated impacts for students with extremely low or high performing peers are substantial. One of the reasons for conducting this analysis was to help shift the focus of policy makers to topics other than peer effects if these estimated effects were found to be small or non-existent. In this case, the answer to the question of the impact of peer ability on one's own achievement for most students is a small amount. When a classroom is organized in such a way that one's peers (on average) have prior achievement levels that are more than one standard deviation above or below the mean, the projected impact on student outcomes can be substantial. This finding complicates the question of whether policy makers and parents should be concerned about the arrangement of students into classrooms. Since it is not possible to arrange students in such a way that all have high performing peers, seeking high performing peers for one set of students denies the benefits of placement with higher performing peers to the remaining students. Based on these findings, assigning students to classrooms randomly would increase equity, but the predicted attainment of high performing students placed in high peer ability classes would be diminished and the predicted performance of low performing students in low peer ability classes would be increased.

For the average student, many other policies have substantially higher associations with student outcomes than peer effects (such as the estimated impacts of first year teacher performance or Nationally Board Certified Teachers in some subjects). Given the distribution of peer effects in this analysis and other analyses of peer effects, any gains for students arranged in classrooms with high performing peers will be offset by lower expected outcomes for classrooms whose average peer prior achievement levels are lowered by the removal of these students into specialized high prior performance level classrooms. Arranging students into classrooms based on peer ability measures can produce differential outcomes for students, but any expected EOC test score increases for some students are likely to generate equity concerns as these increases would likely be offset by lower expected scores for other students. Given the zero-sum nature of the trade-offs with peer ability and the equity goals of public education, policy makers should focus attention on other policy mechanisms besides ‘peer effects.’ Any efficiency gains viable through peer ability grouping alone are minimal and are expected to increase inequity. When considered in concert with the results on tracking, peer effects are more complex.

Consistent with the findings by Zimmer and Toma (2002) and Hanushek, et al. (2003), but with a more detailed micro-level dataset approach, I find no relationship between the variability of prior student performance within a classroom and student EOC test score outcomes in any of the subjects analyzed. This result contradicts findings by Zabel (2008), who found a negative relationship between peer heterogeneity of prior achievement student test score outcomes and Vigdor and Nechyba (2004), who found a similar sized positive relationship between the two. It is possible that the range of prior achievement within a classroom is unrelated to achievement in the current year, but it is also plausible that the different modeling approaches led to different observed relationships. This effect may only be present in classrooms where particular teacher behaviors such as group work or having students present lessons would emphasize this type of knowledge transfer among students. Without additional variables on within classroom behaviors, one must conclude that on average there is no relationship between current year student test performance and the dispersion of prior student test score achievement within classrooms. For policymakers, this conclusion suggests that there is little need for concern with the level of heterogeneity of prior academic achievement within classrooms. More evidence on this relationship is needed.

In terms of tracking results, I find that about three-fourths of high schools engage in some level of tracking behavior based on the observed distribution of students into classrooms compared to a series of randomly generated classroom assignments. Tracking does appear to have a positive relationship with student EOC test score outcomes in Algebra I and Biology but only in schools which utilize comparably high levels of tracking intensity. Biology performance is higher in schools with both moderately and highly tracked classroom assignments while Algebra I test score performance is higher only in schools which utilize the highest levels of tracking. While the tracking intensity in English I was comparable to Biology, no relationship between tracking intensity and student outcomes was observed. Importantly, none of the tracking models showed a negative association between student test score outcomes and the intensity of tracking placements within schools. This does not rule out any negative consequences of tracking on other student outcomes such as graduation rates or social development, but we can conclude that school tracking is not associated with negative test score outcomes in the three subjects studied on average.

While the findings of the included models on tracking indicate that schools which utilize high levels of tracking are correlated with increased EOC outcomes for Algebra I and Biology students, the mechanism for these effects could not be determined in this study. Neither peer effects nor curriculum differences explained the association between higher tracking intensity and EOC outcomes in Algebra I and Biology. Observable teacher characteristics also failed to explain the observed associations. An understanding of mechanisms is not necessary to successfully implement effective policies, but a deeper understanding of the equity of these positive associations may be helpful in generating additional support for tracking. Where tracking intensity is weak, no positive EOC test score outcomes are observed and if other negative consequences are encountered (for example, social stigma for students in certain tracks), a strong argument could be made to discontinue the practice. If intense tracking results in efficiency gains without increasing inequity, arguments for tracking on the basis of student achievement are strengthened. The test of student heterogeneity effects did not reveal a positive relationship between increased diversity of prior performance and student outcomes. Given the size of the coefficients on high intensity tracking in Algebra I and both moderate and high intensity tracking in Biology, a uniform effect across students would imply that students in low performing tracks could actually be equally well off with lower performing peers in an intense tracking environment versus an untracked environment with average peers. Student groups with peers scoring one standard deviation below the mean have predicted decreases in achievement that are smaller than the increases predicted in highly tracked Biology and Algebra I schools. Students in average or high performing tracks would also benefit from the increased track intensity but only in Algebra I and Biology. Testing the interactions between tracking intensity and differing prior student achievement levels is necessary to determine whether or not low performing students are, on average, better off being educated with lower performing peers in an intense tracking environment versus a detracked school. From an efficiency standpoint, more intense tracking is clearly preferred in terms of aggregate student achievement. Other future analysis might explore the relationship between teacher effectiveness (on a metric different from the observable teacher controls included in this study) and tracking intensity. High intensity tracking schools may have more effective teachers, but these teachers may not appear to be different based on measures such as graduate degrees or Praxis test performance. Tracking might be especially effective in certain kinds of schools that are not differentiated based on the variables included in these models such as school leadership or schools with high levels of parental involvement.

Through a series of additional analyses for Algebra I and Biology, I find that curriculum differences between classrooms and peer ability explain only a small portion of the expected difference in student outcomes between schools with little or no tracking and those with high levels of tracking. Observed teacher characteristics explained none of the differences between student outcomes in the different school types. Due to these remaining questions regarding what observed differences might mediate the effects of tracking, the tracking results have provided the most intriguing area for future analyses. Further analysis should also determine if tracking intensity is associated with increasing levels of inequity between students within schools. If tracking effectiveness is explained by schools increasingly devoting resources to advantaged students at the expense of lower performing students, schools are failing in their responsibility to provide equitable opportunities for students, which should be a guiding principle within public school systems.

Broadly, policymakers should consider the findings from this dissertation generally supportive of tracking from an efficiency perspective in high school Algebra I and Biology but only where it is intensely administered (in which intensely administered is defined as schools placing students into classrooms narrowly grouped based on 8^{th} grade EOG test performance). Tracking also appears to function differently across subjects in terms of test score outcomes, and a blanket judgment on the practice either positive or negative on this basis is incorrect. None of the findings in this dissertation suggest negative test score outcomes as a result of a school's tracking intensity. Policy making requires considering the benefits gained through a tracking regime, positive student test score outcomes based on this dissertation, and possible negative consequences such as racial and social stratification (Burris, Wiley, Welner, & Murphy, 2008). The findings from this study suggest decreasing ability tracking within English I classrooms and careful consideration of whether schools track Algebra I and Biology classes in a systematic way based on prior academic achievement. Further study is necessary to precisely determine the academic performance benefits conveyed to students as a result of high intensity school tracking.

While the conclusions above are presented strongly, there are limits to the conclusions that should be drawn from this study. First, while the student-level data used for this analysis provide substantial power to detect effects, it is also limited to a single state in a single year. North Carolina is a diverse state compared to many other U.S. states, but it is always difficult to predict how findings in one place translate to other contexts. It is encouraging to note that control variables included in this study are consistent with findings using data from other parts of the country. The relationships observed in these studies are associations and may or may not be causal connections between the observed variables. If tracking intensity is highly correlated with an unmeasured variable within these models that is also related to student achievement, the effects attributed to tracking intensity may be caused by some other behavior. While the three subjects studied are the most universally end-of-course tested subjects in North Carolina high schools, study results are limited by the selection process of students actually enrolling in these courses. Substantial effort was made to correct for selection bias in the case of Algebra I where this is most problematic, but attempts to extend the results to students unlikely to take certain courses is beyond the scope of this study. These results also only apply to students enrolled in the specified courses in high school grades, and these findings may not extend to younger or older students studying the same subjects.

APPENDIX A

INVERSE MILLS RATIO WEIGHTING

Weighted Algebra I case counts within this analysis are lower than those in English I because a substantial number of students take their Algebra I course work in middle school 8^{th} grade. This differential in course taking generates a selection bias problem since analyses conducted on high school students enrolled in Algebra I excludes all students enrolled in Algebra I during middle school. Ideally, the included analysis estimates the peer and tracking effects of all students had they enrolled in Algebra I during high school. This type of selection effect is addressed via a method developed by James Heckman using the inverse mills ratio to weight cases (Wooldridge 2003). This method requires running a two-stage regression, one to generate the information necessary to calculate weights and then a second regression to test hypotheses.

In the first stage regression, a binary outcome model is estimated which identifies the likelihood of treatment. In this example, the treatment is enrollment in Algebra I during the 9^{th} grade. In order to correct for selection, it is necessary to weight cases higher for students that are most similar to students taking Algebra I in 8^{th} grade but that delayed enrollment in this course until 9^{th} grade.

Table A.1 – Ninth Grade Algebra I Enrollment

Logistic regression					Number of obs =	96358
					LR chi2(18) =	6589.72
					Prob > chi2 =	0
Log likelihood = -63364.174					Pseudo R2 =	0.0494
Enrolled in Alg I in 9th Grade	**Odds Ratio**	**Std. Err.**	**z**	**P>z**	**[95% Conf. Interval]**	
Standardized EOG 8 Reading	1.02	0.01	1.96	0.05	1.00	1.04
Standardized EOG 8 Math	1.12	0.01	10.10	0.00	1.10	1.14
Male	0.88	0.01	-9.28	0.00	0.86	0.90
Black	0.78	0.01	-14.49	0.00	0.75	0.81
Hispanic	1.11	0.04	2.51	0.01	1.02	1.20
Other	0.75	0.02	-9.12	0.00	0.70	0.80
Free Lunch	0.92	0.02	-4.55	0.00	0.89	0.96
Reduced Lunch	1.13	0.03	4.54	0.00	1.07	1.19
Academically or Intellectually Gifted	0.25	0.01	-58.60	0.00	0.24	0.26
Disabled	0.58	0.01	-22.25	0.00	0.55	0.61
PED Less Than High School	0.86	0.02	-5.45	0.00	0.81	0.91
PED Some College	1.12	0.02	6.26	0.00	1.08	1.16
PED College Graduate	0.84	0.02	-9.43	0.00	0.81	0.87
PED Missing	0.80	0.07	-2.71	0.01	0.69	0.94
Overage	0.83	0.05	-3.17	0.00	0.73	0.93
Underage	0.61	0.01	-29.48	0.00	0.59	0.63
Current LEP Student	1.02	0.07	0.22	0.83	0.89	1.16
Former LEP Student	0.77	0.04	-4.46	0.00	0.69	0.87

Table A.1 provides the results from a logistic regression predicting whether or not a student enrolled in English I in grade 9 is also enrolled in Algebra I. Using the same predictor variables that exist in models of EOC achievement generates results suggesting that students scoring higher on 8th grade math and reading EOG exams, Hispanic students, those eligible for reduced lunch, those with parents having some college, and students currently enrolled in Limited English Proficiency programs are over represented in terms of Algebra I enrollment in 9th grade. Students coded as gifted, disabled, or underage are very unlikely to be enrolled in Algebra I in grade 9.

Wooldridge indicates that it is preferable to have a variable included in the selection equation that is not included in the actual regression to test effects (2003, pg. 589). In order to serve as a valid indicator of selection, the variable should be related to the likelihood of enrollment in Algebra I in 9th grade but should not be associated with the EOC score outcomes for Algebra I students. To address this issue, a variable representing the proportion of 8th grade students in feeder middle schools in 2004-05 taking Algebra I is utilized. Feeder middle schools are those supplying students to a particular high school. Feeder middle schools are determined by identifying the source of 8th grade EOG scores for students enrolled in 9th grade English I during the 2004-05 school year. Cases are weighted by the number of students to generate a variable representing the proportion of 8th graders taking Algebra I during the 2004-05 school year (the same year as the current study). This proportion should not affect the Algebra I scores for high school students in 2004-05 but should be correlated with a student taking Algebra I in grade 9. Students coming from middle schools where a higher proportion of students take Algebra I in 8th grade should be more likely to have taken Algebra I during the previous year.

Table A.2 – Algebra I Enrollment with Grade 8 Proportions

Logistic regression					Number of obs =	96358
					LR chi2(19) =	8252.19
					Prob > chi2 =	0
Log likelihood = -62532.938					Pseudo R2 =	0.0619
Enrolled in Algebra I in 9th Grade	**Odds Ratio**	**Std. Err.**	**z**	**P>z**	**[95% Conf. Interval]**	
Standardized EOG 8 Reading	1.03	0.01	2.29	0.02	1.00	1.05
Standardized EOG 8 Math	1.13	0.01	11.08	0.00	1.11	1.16
Male	0.88	0.01	-9.41	0.00	0.86	0.90
Black	0.82	0.01	-11.31	0.00	0.79	0.85
Hispanic	1.10	0.04	2.32	0.02	1.01	1.19
Other	0.76	0.02	-8.62	0.00	0.71	0.81
Free Lunch	0.90	0.02	-6.01	0.00	0.87	0.93
Reduced Lunch	1.10	0.03	3.57	0.00	1.05	1.16
Academically or Intellectually Gifted	0.24	0.01	-59.54	0.00	0.23	0.25
Disabled	0.58	0.01	-22.05	0.00	0.55	0.61
PED Less Than High School	0.85	0.02	-5.71	0.00	0.81	0.90
PED Some College	1.12	0.02	6.33	0.00	1.08	1.16
PED College Graduate	0.89	0.02	-6.33	0.00	0.86	0.92
PED Missing	0.83	0.07	-2.29	0.02	0.71	0.97
Overage	0.83	0.05	-3.07	0.00	0.74	0.93
Underage	0.60	0.01	-29.86	0.00	0.58	0.62
Current LEP Student	1.06	0.07	0.83	0.41	0.93	1.21
Former LEP Student	0.85	0.05	-2.86	0.00	0.76	0.95
Proportion of Students Taking Algebra I in Grade in Student's Grade 8 School	0.25	0.01	-40.26	0.00	0.23	0.27

Table A.2 adds the proportion of students taking Algebra I in 8th grade for feeder middle schools to the regression and reveals a small increase in the pseudo R-squared increases value. The coefficient of the proportion of students enrolled in Algebra I in 8th grade for feeder middle schools is a strong negative predictor of student enrollment in high school Algebra I, indicating that schools with higher Algebra I enrollments have fewer students enrolled in 9th grade Algebra I. Other coefficients are fairly consistent across the two models.

After estimating the models, predicted values of the likelihood of enrolling in Algebra I during high school are obtained, and these values are then used to calculate inverse mills ratios. The inverse mills ratio is the probability density function divided by the cumulative distribution function. These values were calculated using tools within Stata 10.

A number of comparisons were made using the logit and probit approaches for the binary outcomes models, and there is no substantive difference between the model outcomes or the results of the inverse mills ratio calculations. It appears that the number of observations is large enough to overcome any differences that might appear due to differences in the underlying assumptions of the two approaches.

APPENDIX B

MATCHING STUDENT RECORDS TO TEACHERS

Roster data obtained for the research project listed detailed student information including student identification numbers, first and last names, and student birth dates. Teacher information within roster data was more problematic as teacher identification consisted of a 4 character ‘teacher id’ string, an 18 character teacher name field, and local education agency (LEA) and school codes. Teacher names varied widely across schools and commonly omitted first names in favor of prefixes such as Mr. or Mrs. After cleaning data to eliminate punctuation and properly splitting the teacher name field into a presumed first and last name, I attempted matching using an automated process. Potential matches included each unique individual paid within a school and are limited to cases where LEA and school codes match between the roster data and salary data. In most cases, sufficient information was provided to make a positive match with salaried personnel using this automated process.

Using unmatched roster teacher names and teacher identification codes to compare with the names of salaried personnel data by hand resulted in additional matches. Many of these matches were due to common nicknames such as Bill for William, name suffixes which resulted in failures during the automated matching, and omitted first names. Approximately 10 percent of overall matches are coded as hand matches. Overall, across all grade levels and subjects, about 93 percent of unique ‘teacher id’ and teacher name field entries were matched to names included in individual level salary data.

APPENDIX C

DESCRIPTIVE DATA TABLES

Table C.1 – Descriptive Data – English I

English I Descriptive Statistics - Analysis Data						
Individual Level Data	Count	Weighted Count	Mean	Std. Dev.	Min	Max
English I Standardized EOC Score	69158	57257	0.09	0.96	-3.80	3.05
Peer Ability Reading	69158	57257	0.10	0.65	-2.86	2.23
Peer Ability Reading Squared	69158	57257	0.44	0.57	0.00	8.20
Peer Ability Reading Cubed	69158	57257	0.05	1.03	-23.46	11.15
Low-Performing Student * Peer Ability Reading	69158	57257	-0.11	0.34	-2.86	1.54
High-Performing Student * Peer Ability Reading	69158	57257	0.18	0.39	-2.19	1.80
Std Math Score (Grade 8)	69158	57257	0.10	0.96	-3.06	3.23
Std Math Score Squared (Grade 8)	69158	57257	0.94	1.23	0.00	10.41
Std Math Score Cubed (Grade 8)	69158	57257	0.35	3.02	-28.69	33.58
Std Read Score (Grade 8)	69158	57257	0.10	0.94	-3.60	2.91
Std Read Score Squared (Grade 8)	69158	57257	0.89	1.27	0.00	12.97
Std Read Score Cubed (Grade 8)	69158	57257	-0.01	3.24	-46.73	24.58
Male	69158	57257	0.50	0.50	0.00	1.00
Black	69158	57257	0.29	0.46	0.00	1.00
Hispanic	69158	57257	0.04	0.20	0.00	1.00
Other	69158	57257	0.05	0.22	0.00	1.00
White	69158	57257	0.61	0.49	0.00	1.00
Underage Student based on Grade	69158	57257	0.01	0.11	0.00	1.00

Table C.1 – Descriptive Data – English I - Continued

Overage Student based on Grade	69158	57257	0.22	0.42	0.00	1.00
Academically or Intellectually Gifted	69158	57257	0.14	0.35	0.00	1.00
Disabled Student	69158	57257	0.09	0.29	0.00	1.00
Free Lunch	69158	57257	0.27	0.44	0.00	1.00
Reduced Lunch	69158	57257	0.07	0.25	0.00	1.00
Parent Education Less than High School	69158	57257	0.08	0.27	0.00	1.00
Parent Education Some College	69158	57257	0.32	0.46	0.00	1.00
Parent Education College Graduate	69158	57257	0.35	0.48	0.00	1.00
Parent Education Missing	69158	57257	0.01	0.08	0.00	1.00
LEP Services Recipient	69158	57257	0.01	0.10	0.00	1.00
Previous LEP Services Recipient	69158	57257	0.02	0.12	0.00	1.00
Grade 9	69158	57257	0.99	0.11	0.00	1.00
Grade 10	69158	57257	0.01	0.10	0.00	1.00
Grade 11	69158	57257	0.00	0.03	0.00	1.00
Grade 12	69158	57257	0.00	0.02	0.00	1.00
Days Absent	69158	57257	7.22	9.03	0.00	145.00
Low-Performing Student	69158	57257	0.24	0.43	0.00	1.00
High-Performing Student	69158	57257	0.30	0.46	0.00	1.00
Classroom Level Data	Count	Mean	Std. Dev.	Min	Max	
Peer Dispersion (Reading)	4699	0.69	0.19	0.00	1.90	
Number of Students	4699	19.67	6.87	1.00	60.00	
Class Percent Black	4699	32.05	28.45	0.00	100.00	
Class Percent Hispanic	4699	6.09	10.80	0.00	100.00	
Class Percent Other	4699	5.26	9.24	0.00	100.00	
Class Percent White	4699	56.60	30.76	0.00	100.00	
Class Percent Free Lunch	4699	30.09	23.57	0.00	100.00	
Class Percent Reduced Lunch	4699	6.81	7.97	0.00	90.32	

Table C.1 – Descriptive Data – English I - Continued

Advanced Curriculum	4699	0.28	0.45	0.00	1.00	
Remedial Curriculum	4699	0.35	0.48	0.00	1.00	
Advanced Degree	4699	0.31	0.46	0.00	1.00	
Other Licensure	4699	0.09	0.28	0.00	1.00	
National Board Certification	4699	0.11	0.31	0.00	1.00	
Infield	4699	0.74	0.44	0.00	1.00	
Barrons Rating Most or Highly Competive	4699	0.15	0.36	0.00	1.00	
First Year Teacher	4699	0.10	0.29	0.00	1.00	
1 to 2 Years Teacher Experience	4699	0.12	0.33	0.00	1.00	
3 to 4 Years Teacher Experience	4699	0.11	0.31	0.00	1.00	
5 to 9 Years Teacher Experience	4699	0.25	0.43	0.00	1.00	
10 to 14 Years Teacher Experience	4699	0.12	0.33	0.00	1.00	
15 to 19 Years Teacher Experience	4699	0.09	0.28	0.00	1.00	
20 to 24 Years Teacher Experience	4699	0.06	0.24	0.00	1.00	
25 or more Years Teacher Experience	4699	0.16	0.37	0.00	1.00	
Teacher Test Avg (Praxis, NTE, etc.)	4699	0.10	0.65	-2.80	1.97	
Teacher Non-Certified Pay	4699	4759.97	2413.52	300.00	14925.66	
School Level Data	Count	Mean	Std. Dev.	Min	Max	
Reading Track Low	287	0.25	0.43	0.00	1.00	
Reading Track Med	287	0.26	0.44	0.00	1.00	
Reading Track High	287	0.26	0.44	0.00	1.00	
Total Per Pupil Expenditures	287	70.91	11.93	43.47	156.02	
School Size (ADM)	287	1136.74	503.59	84.00	2667.00	
Teacher Turnover	287	20.61	8.36	0.00	72.73	
School Percent Black	287	29.27	23.40	0.00	97.48	

Table C.1 – Descriptive Data – English I - Continued

School Percent Hispanic	287	4.93	4.11	0.00	30.13	
School Percent Other	287	4.68	6.64	0.00	81.90	
School Percent White	287	61.13	25.03	0.90	99.42	
School Percent Free Lunch	287	25.54	13.61	0.00	70.50	
School Percent Reduced Lunch	287	6.49	3.07	0.00	17.98	
Urban	287	0.21	0.41	0.00	1.00	
Rural	287	0.53	0.50	0.00	1.00	

Table C.2 – Descriptive Data – Algebra I

Individual Level Data	Count	Weighted Count	Mean	Std. Dev.	Min	Max
Algebra I Standardized EOC Score	68352	31393	-0.26	0.94	-3.17	3.26
Peer Ability Math	68352	31393	-0.18	0.56	-2.57	1.73
Peer Ability Math Squared	68352	31393	0.34	0.47	0.00	6.63
Peer Ability Math Cubed	68352	31393	-0.13	0.74	-17.06	5.20
Low-Performing Student * Peer Ability Math	68352	31393	-0.18	0.36	-2.57	1.31
High-Performing Student * Peer Ability Math	68352	31393	0.06	0.26	-1.74	1.73
Std Math Score (Grade 8)	68352	31393	-0.24	0.86	-2.71	3.04
Std Math Score Squared (Grade 8)	68352	31393	0.80	1.02	0.00	9.25
Std Math Score Cubed (Grade 8)	68352	31393	-0.44	2.26	-19.84	28.13
Std Read Score (Grade 8)	68352	31393	-0.21	0.93	-3.66	2.91
Std Read Score Squared (Grade 8)	68352	31393	0.90	1.35	0.00	13.36
Std Read Score Cubed (Grade 8)	68352	31393	-0.77	3.42	-48.84	24.58
Male	68352	31393	0.55	0.50	0.00	1.00
Black	68352	31393	0.36	0.48	0.00	1.00
Hispanic	68352	31393	0.06	0.23	0.00	1.00
Other	68352	31393	0.06	0.23	0.00	1.00
White	68352	31393	0.53	0.50	0.00	1.00
Underage Student based on Grade	68352	31393	0.01	0.11	0.00	1.00
Overage Student based on Grade	68352	31393	0.37	0.48	0.00	1.00
Academically or Intellectually Gifted	68352	31393	0.09	0.28	0.00	1.00
Disabled Student	68352	31393	0.16	0.37	0.00	1.00
Free Lunch	68352	31393	0.34	0.47	0.00	1.00
Reduced Lunch	68352	31393	0.07	0.25	0.00	1.00
Parent Education Less than High School	68352	31393	0.11	0.32	0.00	1.00

Table C.2 – Descriptive Data – Algebra I - Continued

Parent Education Some College	68352	31393	0.29	0.45	0.00	1.00
Parent Education College Graduate	68352	31393	0.32	0.46	0.00	1.00
Parent Education Missing	68352	31393	0.01	0.09	0.00	1.00
LEP Services Recipient	68352	31393	0.01	0.10	0.00	1.00
Previous LEP Services Recipient	68352	31393	0.04	0.20	0.00	1.00
Grade 9	68352	31393	0.74	0.44	0.00	1.00
Grade 10	68352	31393	0.20	0.40	0.00	1.00
Grade 11	68352	31393	0.05	0.22	0.00	1.00
Grade 12	68352	31393	0.01	0.10	0.00	1.00
Days Absent	68352	31393	8.05	9.51	0.00	168.00
Low-Performing Student	68352	31393	0.36	0.48	0.00	1.00
High-Performing Student	68352	31393	0.17	0.38	0.00	1.00
Classroom Level Data	Count	Mean	Std. Dev.	Min	Max	
Peer Dispersion (Math)	4770	0.61	0.17	0.00	1.76	
Number of Students	4770	19.17	6.47	1.00	50.00	
Class Percent Black	4770	32.50	27.09	0.00	100.00	
Class Percent Hispanic	4770	6.42	9.03	0.00	100.00	
Class Percent Other	4770	4.75	9.12	0.00	100.00	
Class Percent White	4770	56.33	29.05	0.00	100.00	
Class Percent Free Lunch	4770	32.12	21.42	0.00	100.00	
Class Percent Reduced Lunch	4770	7.60	8.56	0.00	100.00	
Advanced Curriculum	4770	0.09	0.28	0.00	1.00	
Remedial Curriculum	4770	0.41	0.49	0.00	1.00	
Advanced Degree	4770	0.28	0.45	0.00	1.00	
Other Licensure	4770	0.11	0.32	0.00	1.00	
National Board Certification	4770	0.07	0.25	0.00	1.00	
Infield	4770	0.59	0.49	0.00	1.00	
Barrons Rating Most or Highly Competive	4770	0.15	0.36	0.00	1.00	

Table C.2 – Descriptive Data – Algebra I - Continued

First Year Teacher	4770	0.08	0.27	0.00	1.00	
1 to 2 Years Teacher Experience	4770	0.12	0.33	0.00	1.00	
3 to 4 Years Teacher Experience	4770	0.09	0.28	0.00	1.00	
5 to 9 Years Teacher Experience	4770	0.21	0.41	0.00	1.00	
10 to 14 Years Teacher Experience	4770	0.14	0.35	0.00	1.00	
15 to 19 Years Teacher Experience	4770	0.11	0.31	0.00	1.00	
20 to 24 Years Teacher Experience	4770	0.09	0.28	0.00	1.00	
25 or more Years Teacher Experience	4770	0.16	0.37	0.00	1.00	
Teacher Test Avg (Praxis, NTE, etc.)	4770	0.00	0.73	-2.67	2.47	
Teacher Non-Certified Pay	4770	4628.41	2412.28	300.00	25091.63	
School Level Data	Count	Mean	Std. Dev.	Min	Max	
Math Track Low	286	0.30	0.46	0.00	1.00	
Math Track Med	286	0.28	0.45	0.00	1.00	
Math Track High	286	0.22	0.42	0.00	1.00	
Total Per Pupil Expenditures	286	70.77	11.69	43.47	156.02	
School Size (ADM)	286	1142.88	498.40	88.00	2667.00	
Teacher Turnover	286	20.34	7.66	0.00	45.45	
School Percent Black	286	28.59	22.57	0.00	97.48	
School Percent Hispanic	286	4.99	4.17	0.00	30.13	
School Percent Other	286	4.65	6.59	0.00	81.90	
School Percent White	286	61.77	24.49	0.90	99.42	
School Percent Free Lunch	286	25.65	13.39	0.00	70.50	
School Percent Reduced Lunch	286	6.52	3.01	0.00	17.98	
Urban	286	0.20	0.40	0.00	1.00	
Rural	286	0.54	0.50	0.00	1.00	

Table C.3 – Descriptive Data – Biology

Individual Level Data	Count	Weighted Count	Mean	Std. Dev.	Min	Max
Biology Standardized EOC Score	53098	43101	0.04	0.97	-3.72	3.44
Peer Ability Reading	53098	43101	0.15	0.55	-2.88	1.71
Peer Ability Reading Squared	53098	43101	0.33	0.42	0.00	8.31
Peer Ability Reading Cubed	53098	43101	0.13	0.64	-23.97	4.96
Low-Performing Student * Peer Ability Reading	53098	43101	-0.06	0.23	-2.88	1.39
High-Performing Student * Peer Ability Reading	53098	43101	0.17	0.36	-1.32	1.71
Std Math Score (Grade 8)	53098	43101	0.15	0.94	-3.01	3.56
Std Math Score Squared (Grade 8)	53098	43101	0.90	1.21	0.00	12.68
Std Math Score Cubed (Grade 8)	53098	43101	0.49	2.96	-27.18	45.14
Std Read Score (Grade 8)	53098	43101	0.14	0.91	-3.66	2.91
Std Read Score Squared (Grade 8)	53098	43101	0.85	1.21	0.00	13.36
Std Read Score Cubed (Grade 8)	53098	43101	0.13	3.01	-48.84	24.58
Male	53098	43101	0.48	0.50	0.00	1.00
Black	53098	43101	0.29	0.45	0.00	1.00
Hispanic	53098	43101	0.03	0.18	0.00	1.00
Other	53098	43101	0.05	0.21	0.00	1.00
White	53098	43101	0.63	0.48	0.00	1.00
Underage Student based on Grade	53098	43101	0.01	0.11	0.00	1.00
Overage Student based on Grade	53098	43101	0.21	0.41	0.00	1.00
Academically or Intellectually Gifted	53098	43101	0.13	0.34	0.00	1.00
Disabled Student	53098	43101	0.07	0.26	0.00	1.00
Free Lunch	53098	43101	0.23	0.42	0.00	1.00
Reduced Lunch	53098	43101	0.06	0.24	0.00	1.00
Parent Education Less than High School	53098	43101	0.07	0.25	0.00	1.00
Parent Education Some College	53098	43101	0.35	0.48	0.00	1.00

Table C.3 – Descriptive Data – Biology – Continued

Parent Education College Graduate	53098	43101	0.36	0.48	0.00	1.00
Parent Education Missing	53098	43101	0.00	0.05	0.00	1.00
LEP Services Recipient	53098	43101	0.01	0.09	0.00	1.00
Previous LEP Services Recipient	53098	43101	0.01	0.10	0.00	1.00
Grade 9	53098	43101	0.22	0.42	0.00	1.00
Grade 10	53098	43101	0.65	0.48	0.00	1.00
Grade 11	53098	43101	0.11	0.31	0.00	1.00
Grade 12	53098	43101	0.02	0.14	0.00	1.00
Days Absent	53098	43101	7.17	8.26	0.00	122.00
Low-Performing Student	53098	43101	0.21	0.41	0.00	1.00
High-Performing Student	53098	43101	0.31	0.46	0.00	1.00
Classroom Level Data	Count	Mean	Std. Dev.	Min	Max	
Peer Dispersion (Reading)	3472	0.73	0.19	0.00	2.02	
Number of Students	3472	20.66	6.17	2.00	35.00	
Class Percent Black	3472	30.61	27.43	0.00	100.00	
Class Percent Hispanic	3472	4.76	7.00	0.00	100.00	
Class Percent Other	3472	5.21	8.50	0.00	100.00	
Class Percent White	3472	59.41	29.49	0.00	100.00	
Class Percent Free Lunch	3472	25.29	20.53	0.00	100.00	
Class Percent Reduced Lunch	3472	6.36	7.26	0.00	60.00	
Advanced Curriculum	3472	0.27	0.45	0.00	1.00	
Remedial Curriculum	3472	0.35	0.48	0.00	1.00	
Advanced Degree	3472	0.35	0.48	0.00	1.00	
Other Licensure	3472	0.09	0.29	0.00	1.00	
National Board Certification	3472	0.11	0.31	0.00	1.00	
Infield	3472	0.85	0.35	0.00	1.00	
Barrons Rating Most or Highly Competive	3472	0.21	0.40	0.00	1.00	
First Year Teacher	3472	0.07	0.26	0.00	1.00	
1 to 2 Years Teacher Experience	3472	0.10	0.31	0.00	1.00	
3 to 4 Years Teacher Experience	3472	0.13	0.33	0.00	1.00	

Table C.3 – Descriptive Data – Biology – Continued

5 to 9 Years Teacher Experience	3472	0.22	0.42	0.00	1.00	
10 to 14 Years Teacher Experience	3472	0.13	0.34	0.00	1.00	
15 to 19 Years Teacher Experience	3472	0.11	0.31	0.00	1.00	
20 to 24 Years Teacher Experience	3472	0.09	0.28	0.00	1.00	
25 or more Years Teacher Experience	3472	0.15	0.35	0.00	1.00	
Teacher Test Avg (Praxis, NTE, etc.)	3472	0.17	0.67	-1.77	2.25	
Teacher Non-Certified Pay	3472	4958.29	2567.43	245.45	21827.12	
School Level Data	Count	Mean	Std. Dev.	Min	Max	
Reading Track Low	258	0.28	0.45	0.00	1.00	
Reading Track Med	258	0.28	0.45	0.00	1.00	
Reading Track High	258	0.24	0.43	0.00	1.00	
Total Per Pupil Expenditures	258	72.18	13.45	48.73	188.56	
School Size (ADM)	258	1101.59	483.39	84.00	2667.00	
Teacher Turnover	258	20.82	8.60	0.00	72.73	
School Percent Black	258	28.99	22.48	0.00	93.90	
School Percent Hispanic	258	5.03	4.29	0.00	30.13	
School Percent Other	258	4.73	6.89	0.00	81.90	
School Percent White	258	61.25	24.61	0.90	99.42	
School Percent Free Lunch	258	25.98	13.61	0.00	70.50	
School Percent Reduced Lunch	258	6.61	3.05	0.00	17.98	
Urban	258	0.21	0.41	0.00	1.00	
Rural	258	0.54	0.50	0.00	1.00	

REFERENCES

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.

Angrist, J. D., & Lavy, V. (1999). Using Maimonides' Rule to Estimate The Effect of Class Size on Scholastic Achievement*. *Quarterly Journal of Economics, 114*(2), 533-575.

Betts, J. R., & Shkolnik, J. L. (2000). The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review, 19*(1), 1.

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.

Burke, M. A., & Sass, T. R. (2008). *Classroom Peer Effects and Student Achievement. Working Paper 08-5*. Boston, MA: Federal Reserve Bank of Boston.

Bryk, A. S., & Raudenbush, S. W. (1988). Toward a More Appropriate Conceptualization of Research on School Effects: A Three-Level Hierarchical Linear Model. *American Journal of Education, 97*(1), 65.

Burris, C. C., Wiley, E., Welner, K. G., & Murphy, J. (2008). Accountability, Rigor, and Detracking: Achievement Effects of Embracing a Challenging Curriculum As a Universal Good for All Students. *Teachers College Record, 110*(3).

Carbonaro, W. (2005). Tracking, Students' Effort, and Academic Achievement. *Sociology of Education, 78*(1), 27-49.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). *Teacher-Student Matching and the Assessment of Teacher Effectiveness*: SSRN.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., et al. (1966). *Equality of Educational Opportunity*. Washington, DC: U.S. Government Printing Office.

Cook, T. D., Deng, Y., & Morgano, E. (2007). Friendship Influences During Early Adolescence: The Special Role of Friends' Grade Point Average. *Journal of Research on Adolescence, 17*(2), 325-356.

Ehrenberg, R. G., & Brewer, D. J. (1994). Do school and teacher characteristics matter? Evidence from High School and Beyond. *Economics of Education Review, 13*(1), 1.

Figlio, D. N. (1999). Functional form and the estimated effects of school resources. *Economics of Education Review, 18*(2), 241.

Figlio, D. N., & Page, M. E. (2002). School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality? *Journal of Urban Economics, 51*(3), 497-514.

Gamoran, A. (1992). The Variable Effects of High School Tracking. *American Sociological Review, 57*(6), 812-828.

Goldhaber, D., & Anthony, E. (2005). Can Teacher Quality Be Effectively Assessed? National Board Certification as a Signal of Effective Teaching. The Urban Institute.

Goldhaber, D. D., & Brewer, D. J. (1996). *Evaluating the Effect of Teacher Degree Level on Educational Performance*. Rockville, MD.: Westat, Inc.

Hallinan, M. T. (1990). The Effects of Ability Grouping in Secondary Schools: A Response to Slavin's Best-Evidence Synthesis. *Review of Educational Research, 60*(3), 501.

Hallinan, M. T. (1994). Tracking: From Theory to Practice. *Sociology of Education, 67*(2), 79-84.

Hanushek, E. (1997). Assessing the Effects of School Resources on Student Performance : An Update. *Educational Evaluation and Policy Analysis, 19*(2).

Hanushek, E. A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature, 24*(3), 36.

Hanushek, E. A. (2003). The Failure of Input-Based Schooling Policies. *Economic Journal, 113*(485), F64.

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics, 18*(5), 527-544.

Henderson, V., Mieszkowski, P., & Sauvageau, Y. (1978). Peer group effects and educational production functions. *Journal of Public Economics, 10*(1), 97.

Henry, G. T., & Rickman, D. K. (2007). Do peers influence children's skill development in preschool? *Economics Of Education Review, 26*(1), 100-112.

High School Graduation Requirements. Retrieved from http://www.ncpublicschools.org/docs/curriculum/home/graduationrequirements.pdf.

Hoffer, T. B. (1992). Middle School Ability Grouping and Student Achievement in Science and Mathematics. *Educational Evaluation and Policy Analysis, 14*(3), 205.

Honors Course Implementation Guide (2005). Raleigh, NC: Public Schools of North Carolina.

Hox, J. (2002). *Multilevel Analysis: Techniques and Applications*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Hoxby, C. (2000). *Peer Effects in the Classroom: Learning from Gender and Race Variation*.

Hoxby, C. M., & Weingarth, G. (2006). *Taking Race Out of the Equation: School reassignment and the structure of peer effects*. Harvard University.

Innovations in Education: Creating Strong District School Choice Programs (2004). Washington, D.C.: U.S. Department of Education, Office of Innovation and Improvement.

Kim, J.-S. (2009). Multilevel Analysis: An Overview and Some Contemporary Issues. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE Handbook of Quantitative Methods in Psychology*. Thousand Oaks, California: SAGE Publications Inc.

Kiuru, N., Aunola, K., Vuori, J., & Nurmi, J.-E. (2007). The Role of Peer Groups in Adolescents' Educational Expectations and Adjustment. *Journal of Youth and Adolescence, 36*(8), 995-1009.

Mark, M. M., Henry, G. T., and Julnes, G. (2000). *Evaluation: an integrated framework for understanding, guiding, and improving policies and programs.* San Francisco: Jossey-Bass.

Monk, D. H. (1992). Education Productivity Research: An Update and Assessment of Its Role in Education Finance Reform. *Educational Evaluation And Policy Analysis, 14*(4), 307-332.

No Child Left Behind Act of 2001, 1425(2001).

Oakes, J. (1994). More than Misapplied Technology: A Normative and Political Response to Hallinan on *Tracking*. Exchange. *Sociology of Education, 67*(2), 84-89.

Outline of the Course Coding Structure For NC Public Schools 2009-2010 (2008). Retrieved June 27, 2010. from http://www.ncpublicschools.org/docs/curriculum-instruction/home/2009-10coursecodes.pdf.

Parents Involved in Community Schools v. Seattle School District No. 1 (U.S. Supreme Court 2007).

Quick Guide to Using the NELS:88/2000 Data (2002). Washington, D.C.: National Center for Education Statistics.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models*. Thousand Oaks, CA: Sage Publications, Inc.

Rees, D. I., Brewer, D. J., & Argys, L. M. (1999). How should we measure the effect of ability grouping on student performance? *Economics of Education Review, 19*(1), 17.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica, 73*(2), 417-458.

Rothstein, J. (2009). Student Sorting and Bias in Value-Added Estimation: Selection on Observables and Unobservables. *Education Finance and Policy, 4*(4), 537-571.

Sander, W. (1999). Endogenous expenditures and student achievement. *Economics Letters, 64*(2), 223.

Slavin, R. E. (1990). Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis. *Review of Educational Research, 60*(3), 471.

Stiefel, L., Schwartz, A. E., & Ellen, I. G. (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.

Summers, A. A., & Wolfe, B. L. (1977). Do Schools Make a Difference? *The American Economic Review, 67*(4), 639.

Todd, P. E., & Wolpin, K. I. (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement*. *The Economic Journal, 113*(485 %R doi:10.1111/1468-0297.00097), F3-F33.

Vigdor, J., & Nechyba, T. (2004). Peer Effects in North Carolina Public Schools. Duke University and NBER.

Wooldridge, J. M. (2003). *Introductory Econometrics: A Modern Approach, 2e.* Mason, OH: South-Western.

Zabel, J. E. (2008). The Impact of Peer Effects on Student Outcomes in New York City Public Schools. *Education Finance and Policy, 3*(2), 197-249.

Zelman Superintendent of Public Instruction of Ohio, et al. v. Simmons-Harris et al., 536 U.S. 639 (U.S. Supreme Court 2002).

Zimmer, R. (2003). A new twist in the educational tracking debate. *Economics of Education Review, 22*(3), 307.

Zimmer, R. W., & Toma, E. F. (2000). Peer effects in private and public schools across countries. *Journal of Policy Analysis and Management, 19*(1), 75-92.