THE UNIVERSITY OF CHICAGO

CAN YOU "WORK YOUR WAY UP?" – ABILITY GROUPING AND THE DEVELOPMENT OF ACADEMIC ENGAGEMENT

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF SOCIAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF SOCIOLOGY

BY

MARSHALL JEAN

CHICAGO, ILLINOIS

AUGUST 2016

Table of Contents

List of Tables

List of Figures

Acknowledgements

This research would not have been possible without the support of both key individuals and institutions. First, I would like to thank my committee members. Steve Raudenbush, my advisor and mentor for nearly a decade, has been instrumental in my development as a scholar and educator. Guanglei Hong was always available to help think through the finer points of methodology and provide expertise on grouping in the primary grades. Sara Stoelinga kindly stepped up to join my committee on short notice, and provided detailed and valuable critiques to help refine the story told in the following chapters. Charles Bidwell provided some of my earliest training as a sociologist of education, and his input shaped the research questions and framing of this dissertation. I would also like to thank my wife, Ashley, for endless emotional support and proofreading.

The groundwork for this project was initially supported by a predoctoral fellowship provided by the Institute of Education Sciences. The analysis and write up was supported by a grant from the Successful Pathways from School to Work Project, an initiative of the Hymen Milgrom Supporting Organization, The University of Chicago's Committee on Education, and the Division of Social Sciences. However, I am the sole analyst and author, so all opinions and errors remain my own.

Abstract

Schools are not only tasked with communicating academic knowledge to students, but also preparing and training them to assume adult roles, including participation in the work force. Educators attempt to foster engagement with learning and cultivate student behaviors such as punctual attendance and preparation for classes, attention to lectures and directions, and timely completion of assigned tasks. These learning behaviors are crucial to students' success in school and their preparation for the labor market. However, the widespread practice of grouping students by perceived ability means that students in the same school, or even the same classroom, can have very different training experiences. While there is a large literature on how ability grouping can affect student achievement, we have comparatively little empirical evidence on how it affects the development of students' academic engagement. I consider two different types of grouping – within-class grouping in the early elementary grades, and between-class grouping (tracking) during high school – using data from the Early Childhood Longitudinal Study Kindergarten Cohort of 2002 and the National Education Longitudinal Study of 1988. I describe predictors of students' assignments to ability groups and examine opportunities for advancement. I also use propensity score stratification to estimate the effects of group assignments on math and reading achievement as well as students' academic engagement through its attendant learning behaviors.

Findings indicate that there are substantial opportunities for students to move up to higher groups as they progress through school and students can improve their group assignments by demonstrating positive academic engagement. However, group assignments do tend to "stick" to students over time, even after controlling for prior achievement, learning behaviors, and other observable characteristics. Moreover, assignment to a higher group improves students' math and reading achievement outcomes and improves students' academic engagement, while assignment to a lower group depresses both of these, making upward mobility more difficult. These effects diminish

in size as students age. Following high school, students who graduated from an academic track were more likely to attend traditional four-year colleges (though less likely to attend other postsecondary education programs), and differences in labor market outcomes were modest.

Chapter 1 – Theoretical and Empirical Background

Introduction

While schools are primarily tasked with teaching students particular academic skills and content knowledge, they also play an important role in preparing students for adult life. Other institutions such as churches, community groups, or clubs shape students' perceptions and behaviors, but students spend far more time in schools than in any other context, excepting their own families. More importantly, the organization of schools and the content of instruction can be changed though public policy. Sociologists such as Talcott Parsons (1959) and Robert Dreeben (1968) described how schools prepare students for life outside of the households into which they are born. Students learn to direct their attention to lessons, to balance their own initiative with the demands of authority figures, and to complete assigned tasks on time. These skills, which I call learning behaviors in the context of schools, represent many of the same skills and dispositions that students will need to succeed in the labor market as adults. In this way, schools serve as the training grounds to prepare students for the workplace.

Learning behaviors are visible, purposeful actions directed toward learning. These include, for example, regularly attending class (on time and with all necessary materials), paying attention to lectures, participation in classroom discussions and activities, and spending time outside of school to complete homework. They are related to, but distinct from, academic mindsets such as motivation to learn or interest in the material. Rather, they are the "outward signs that a student is engaged and putting forth effort to learn" (Farrington et al., 2012). They are central to the development of academic achievement, as even very intelligent students cannot master all skills and material presented to them through passive observation alone. Teachers often rely on observing these behaviors in order to help gauge how well students are learning.

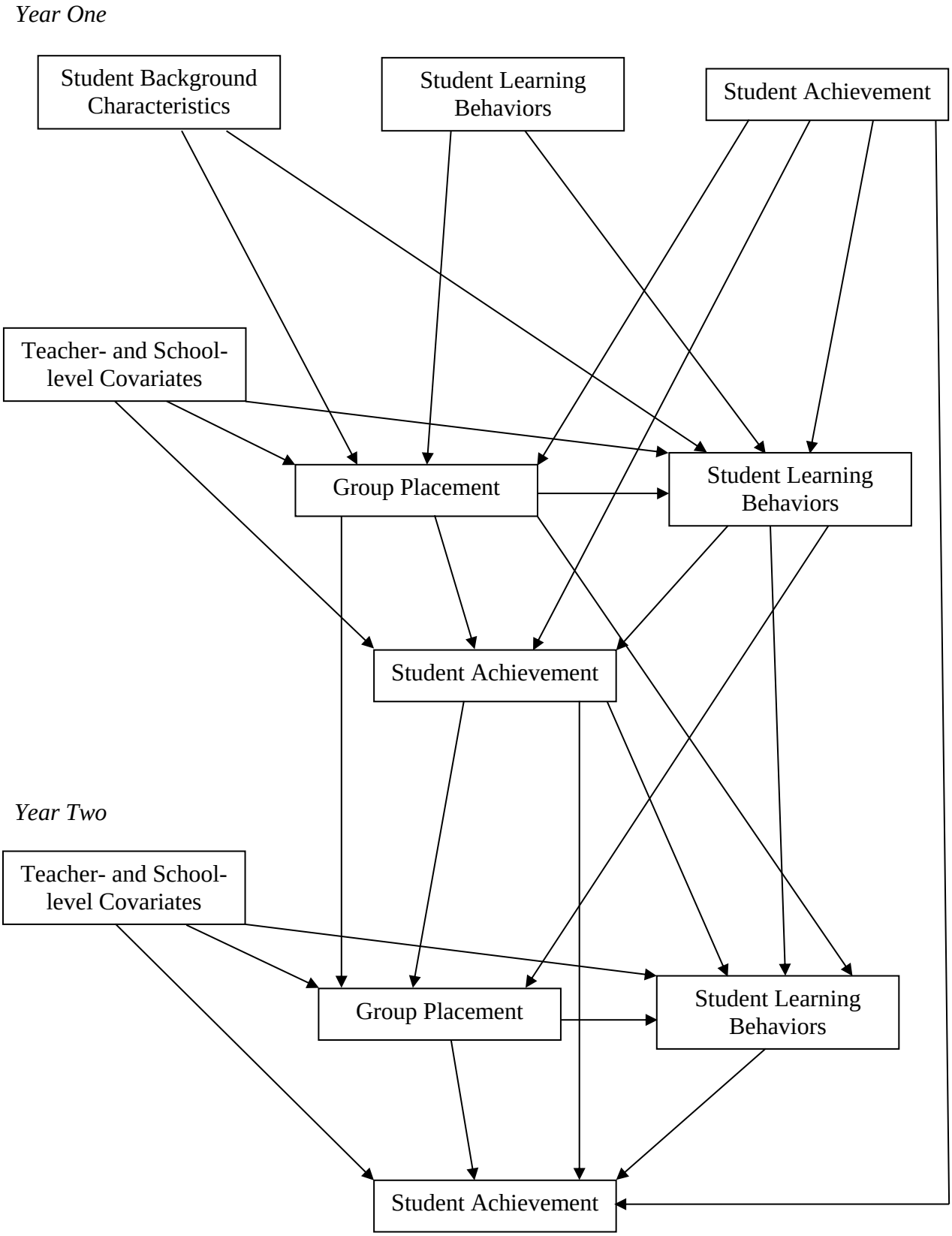
Learning behaviors are also different from academic perseverance and "grit," although behaviors may change depending on how resilient students are when faced with difficulties.

Schools encourage these learning behaviors by setting formal rules and informal expectations, and then rewarding or sanctioning student behavior. Behaviors such as punctuality, preparation for class, attention to lectures and discussion, and timely assignment completion are rewarded; behaviors such as talking out of turn, tardiness or truancy, disrupting the class, or showing a lack of effort are punished. However, the way students experience these sanctions can vary substantially, even within the same classroom. The practice of grouping students according to perceived ability means that students experience different instructional regimes and interactions with peers and teachers. As early as the first week of kindergarten, students are often divided into groups for differentiated instruction, most often for reading (Eder and Felmee, 1984; Weinstein, 1976). About two-thirds of kindergarten teachers use ability grouping, and this proportion increases through third grade (Chapter 2 analysis). At the high school level, students typically select their own course schedule with the aid of parents and counselors (Delany, 1991). They can choose from different levels of core classes (for example, ninth graders may be able to choose from pre-algebra, algebra, geometry, or business math), but these choices are often limited by structural requirements such as minimum test scores, teacher or counselor recommendation, or pre-requisite coursework. Each of these is based on assumptions regarding requisite levels of ability or preparation, and act as signals directing students toward or away from particular courses of study.

Because the practice of ability grouping has been both extremely common and controversial for many decades, a large body of research exists examining the relationship between grouping practices and student academic achievement. However, we know very little

about how grouping affects non-academic outcomes, particularly learning behaviors (Tach and Farkas, 2006) and students' experiences in the labor market following high school (Moller and Sterns, 2012). Moreover, much of the existing literature focuses only on grouping as a structural condition of which students are passive recipients, rather than actors that influence and react to their experiences in the classroom. It is my intention to explore how students' own inputs into the learning process – namely their learning behaviors – are both a cause and consequences of their structural locations. If mobility between group levels is possible and at least somewhat common, then one can imagine an iterative process as students move through school, in which students are assigned to groups (on the basis of prior academic outcomes, learning behaviors, and other information), these group assignments affect future academic outcomes and learning behaviors, and these in turn affect future group assignments. Consider the following conceptual model:

Figure 1 – Conceptual Model



Research on mobility between group levels is somewhat scarce, but there is reason to believe that it is possible and opportunities for mobility may have increased in recent decades. At the primary school level, the modal pattern described in classrooms that use grouping is that teachers will form groups early in the year, generally within the first month or so. Teachers will then tinker with group assignments, occasionally moving individual students up or down during the year. Teachers can also split or combine groups. During-year mobility estimates vary considerably, with earlier studies usually reporting lower mobility rates. Rist (1970) described mobility as very rare and upward mobility as virtually impossible for disadvantaged students. Groff (1962) found 16% of students changed reading groups during the first semester of a school year, and Hawkins (1966) observed rates of about 9% during a similar period. Analysis of data from the 1960's revealed group transfers occurring in 12 out of 15 classrooms, with upward mobility rates of 15% and downward rates of 14% (Dreeben, 1984). Weinstein (1976) found approximately 36% of pupils were reassigned between October and January, noting that mobility patterns varied substantially by classroom, with patterns of upward, downward, and no mobility. Eder (1983) documented fourteen (mostly upward) moves among 22 students over a school year.

However, there are structural factors limiting the amounts of group mobility that is practical. Physical space and the need for the teacher to divide his or her attention among groups limit the number of groups that can be formed. Smaller groups are easier to manage, so it is less likely that teachers will allow groups to grow to accommodate any number of students (Dreeben, 1984; Dreeben and Barr, 1988; Eder, 1983). Parents may resist the movement of their children into lower reading groups. Norms of fairness and equal access to educational resources discourage teachers from spending more time with some pupils while ignoring others (Hallinan and Sørenson, 1983). Educators must therefore balance organizational concerns with the

instructional needs of the students. This might involve some tolerance of mismatch between students' levels of preparedness and their group assignments (Delany, 1991; Eder, 1983). Clark-Ibanez (2005) describes a teacher "promoting" three students in a lower group by providing new spelling word lists used by the middle group, but without changing the students' group affiliation. But these structural concerns are less constraining as students transition between school years. With complete reorganization of classes and a new teacher evaluating individual students' achievement and efforts, mobility may be easier. Longitudinal studies that track students' within-class group assignments over more than a single school year are rare, but there is reason to believe that group assignments have a certain amount of "stickiness" that can persist even as students progress through school. Research from the United Kingdom (Kerckhoff, 1993) and United States (Gamoran, 1989) indicate that group assignments in earlier years influence later assignments even after controlling for realized student achievement. Nonetheless, group homogeneity in terms of prior achievement tends to increase in later grades, suggesting that students are indeed re-assessed between grades. Opportunities for mobility may decline as students age, especially once they enter high schools with more formal grouping arrangements and requirements for entry into higher-level classes, such as minimum standardized test scores, prerequisite classes, or teacher recommendations. To my knowledge, this project is the first to use national-scale data to consider group mobility (both within and between school years) as students move through elementary school, and the first to compare mobility rates early and late in students' educational careers.

Possibilities for mobility and the patterns of observed upward and downward mobility might also affect student learning behaviors. Building on work by Blauner (1964), who found that workers were less likely to be alienated when organizations are characterized by fine

hierarchical distinctions and greater promotion, Richer (1976) suggests that perceived opportunities for mobility is a critical component in determining how students respond to group assignments. As perceived opportunities for upward mobility increase, it is more likely that students assigned to lower groups will respond by aspiring to become members of higher groups and less likely that they will be discouraged by their assignment. Perceptions of mobility opportunities by students can be affected by several factors, including the actual opportunities for mobility, experiences in different group levels in different contexts (such as different group assignments across different subjects or grades), and correlations between group assignments and ascribed student characteristics.

Of course, not all students will respond to group assignments or possibilities for mobility the same way. Some students might prefer to remain in groups where they have friends rather than move to new ones. Others may be sensitive to the possibility of being seeing to struggle with more demanding material in front of peers. Some students may simply believe that they are incapable of performing at a higher level even if they exert more effort. At around ages ten to twelve, children begin to distinguish between effort and ability as separate factors, and begin to form distinct beliefs about how one affects the other. Students who believe that academic ability is stable and unchangeable are perhaps less likely to be enticed by the possibility of upward mobility. Students with a growth mindset, who believe that abilities can be improved through practice, may be more motivated. These two mindsets can also shape how experiences in grouped instruction can influence future learning behaviors. Students with a fixed mindset that fail when confronted with challenging material are more likely to avoid challenges in the future, and are less likely to work hard to overcome future challenges (Dweck, 2001).

Research Questions

It is my intention to explore this iterative process of grouping and regrouping, describe possibilities for group mobility, and better understand how group assignments can affect student learning behaviors. The central motivating question for this research is: "As students progress through school, can they work their way up into higher positions and more demanding learning environments?" Specifically:

1. What student background, academic, and learning behaviors predict students' group assignments?
2. How much mobility between group levels occurs, and what academic and behavioral characteristics predict this mobility?
3. How does assignment to a high or low ability group affect academic outcomes and subsequent learning behaviors?
4. How do group assignments in high school affect success in the labor market?

This project includes two sets of analyses, detailed in the following chapters. First, I examine within-class grouping at the primary school level using data from the Early Childhood Longitudinal Study - Kindergarten Cohort (ECLS-K). Initiated by the National Center for Education Statistics (NCES), the ECLS-K collected information from a nationally representative sample of kindergartners, their parents, teachers, and schools across the United States. It followed the same children from kindergarten through the 8th grade. Information was collected in the fall and the spring of kindergarten (1998-99), the fall and spring of 1st grade (1999-2000), the spring of 3rd grade (2002), the spring of 5th grade (2004), and the spring of 8th grade (2007). Children, their families, their teachers, and their schools provided information on children's cognitive, social, and emotional development. Information on children's home environment,

home educational activities, school environment, classroom environment, classroom curriculum, and teacher qualifications was also collected. Data on reading and math grouping practices, and the group placement of individual students, was collected during the 1st and 3rd grade spring waves and these waves form the focal point of my analysis. Information collected during kindergarten supplies pre-treatment covariates.

Second, I consider between-class grouping at the secondary level and draw on data from the National Education Longitudinal Study of 1988 (NELS). A nationally-representative sample of eighth-graders were first surveyed in the spring of 1988. A sample of these respondents were then resurveyed through four follow-ups in 1990, 1992, 1994, and 2000. On the questionnaire, students reported on a range of topics including: school, work, and home experiences; educational resources and support; the role in education of their parents and peers; neighborhood characteristics; educational and occupational aspirations; and other student perceptions. For the three in-school waves of data collection (when most were eighth-graders, sophomores, or seniors), achievement tests in reading, mathematics, and other subjects were administered in addition to the questionnaire. Teachers were surveyed and asked questions about themselves and their perceptions of sampled students. Administrators provided information about schools the sampled students attended. Subsequent waves collected data on postsecondary education and labor market outcomes after students left high school.

To address the heavy selection bias endemic to group assignments, I employ inverse probability of treatment weighting (IPTW). Additional details on methodology are described in the following chapters. It is important to note that there are two related but distinct counterfactuals to consider regarding individual outcomes of students that undergo grouped instruction. The first considers what would have happened if the student had not been sorted into

a group and instead had been instructed in a heterogeneous peer environment. The second is that, given a student is grouped, what would have happened if the student has been assigned to a *different* group? I focus on this second counterfactual question.

In this chapter, I review the empirical literature on how grouping practices affect academic outcomes, discuss theoretical literature on the development of student learning behaviors, and conclude with research questions to be investigated that connect the two fields. First, however, it is helpful to understand how the most common grouping practices in the United States originated and was shaped historically, in order to gain insight into how they may function in practice.

A Brief History

During the early nineteenth century, students of all ages and abilities were typically taught together, and the proverbial "one room schoolhouse" was the modal institution of American education. In 1848, the first school to separate students by grade appeared, by 1860 most larger cities had graded schools, and by 1870 the age-graded school was firmly established as the dominant model of school organization (Kulik and Kulik, 1982). Most children attended local primary schools, but outside of major cities only those from affluent or professional families continued to private academies to prepare for university admission. As mandatory education laws became more commonplace in cities toward the end of the century, enrollment often outpaced the human and physical resources dedicated to education. Class sizes in many places exceeded one hundred, and instruction relied heavily on recitation and memorization with little differentiated instruction (Tyack, 1974). With the beginning of the twentieth century, population growth (which included a substantial amount of immigration), demanded increasing

coordination and standardization of public education. Between 1880 and 1912, school enrollment across the nation increased by over seven hundred percent, from about 200,000 to over 1.5 million. By 1920, nearly 60 percent of fourteen- to seventeen-year-olds were enrolled in high schools (Oakes, 2005). Schools struggled with the dual challenges of dramatic enrollment increases and increased heterogeneity among student social and economic backgrounds. In particular, the urban schools experienced a massive influx of children of recent immigrants. A large study in 1909 by the US Immigration commission found that 57.8 percent of the children in the schools of thirty-seven of the nation's largest cities were of foreign-born parentage (Cremin, 1961). Schooling was increasingly viewed as a way to "Americanize" these youth and socialize them to adopt mainstream views of hygiene, punctuality, deference to authority, and a Protestant religious outlook (Cremin, 1961; Tyack, 1974).

The eventual organizational response was the formation of the comprehensive high school, in which students were sorted for differential instruction that depended on students' likely destinations after graduation. Typically high schools offered an academic track to prepare a select group of students for college admission and a general track to provide basic instruction for most others. Later, many schools began offering vocational instruction that emphasized training for skilled manual labor. This sorting was aided by the advent of standardized intelligence tests. Between 1905 and 1908, psychologists Alfred Binet and Theodore Simon developed idea of a scale of intelligence that could be assessed by a series of questions of graded difficulty, and other researchers created elaborated versions of these tests. These included an adaptation by Lewis Terman of Stanford in 1916 that found widespread adoption in the United States. By 1918, well over a hundred different standardized tests designed to measure achievement in principle elementary and secondary school subjects were in circulation (Cremin,

1961). These tests provided both a means for sorting students into these programs and legitimated this form of organization as meritocratic and scientific. The structure of tracked instruction reflected the widespread view that measured intelligence quotients represented the unidimensional intellectual ability of students and that it was a stable, biologically-determined characteristic. Thus, track assignments determined the level of instruction across all academic subjects with differentiated levels, and mobility between tracks was low (Oakes, 2005, Lucas, 1999).

In the 1960's, these rigid structures began to break down. In 1965, when the Educational Testing Services collected data for the Academic Growth Study, researchers were able to determine the track position of 93% of high school students (Alexander, Cook, and McDill, 1978). But by 1993, only 15 percent of schools had traditional tracking programs, and 71 percent indicated that while they offered differentiated courses, any student could enroll in any level course provided they had taken the prerequisites (Carey, Farris, and Carpenter, 1994). This "unremarked revolution" in formal grouping policies was not a result of any single initiative, but a confluence of changes in the academic world. The theory that intelligence is a single, stable, unitary concept became contested. More importantly, the civil rights movement galvanized support for social justice and sensitized school administrators to issues of educational equity. Because poor and minority students were disproportionately assigned to low tracks, the politics of ability grouping became intertwined with those of desegregation. In many districts, within-school segregation replaced between-school segregation under the guise of ability grouping (Tyson, 2011). In 1967, the court decision in *Hobson v. Hansen* abolished tracking in Washington, DC and provided precedent for the legal challenge of tracking policies in a number of districts (Lucas, 1999). However, this does not mean that student sorting and curricular

differentiation has disappeared or even substantially reduced, but rather the complexity of curricular ladders and assignment mechanisms have increased. More recently, student choice, teacher recommendations, and subject-specific standardized tests have replaced intelligence testing as the primary means for assigning students to courses (Useem, 1992). In the 1980's and 1990's, reformers began to advocate for the elimination of ability grouping in schools, a policy position that was adopted in many districts and some states, such as California and Massachusetts (Loveless, 1999). These reforms were in turn contested among parents, school personnel, and researchers, and the debate over the uses of ability grouping continues to the present.

Structural Theories of Ability Grouping

Before proceeding with research findings, it is prudent to review some of the influential structural theories of tracking and ability grouping that have informed empirical research. These theories have generally been developed to explain more visible, formal practices of ability grouping such as differentiated courses at the secondary level, but can also be adapted to explain informal grouping practices such as within-class grouping at the elementary level. In an early essay, Turner (1960) describes the grouping practices in United States schools as a manifestation of a regime of "contest mobility" in which students compete to win elite status through their own talent and effort. Formal separation and differentiation of academic classes is delayed until secondary school, and there all students have the possibility to advance into the highest tracks. Competition is meritocratic and upward mobility is encouraged. Less successful individuals are even provided remediation to help them advance. Also, success is never ensured, and high-level students must continue to excel lest they be displaced by their peers. This is contrasted to a

system of "sponsored mobility" in the United Kingdom, where elite status is conferred on young people by others, the streaming of students into different classes begins early, and entrance to grammar schools is limited to those that can pass a battery of examinations (Kerckhoff, 1993; Turner, 1960).

Rosenbaum (1976) drew on Turner's theory of contest mobility in American schools but refines it into a tournament model. He argued that the contest model is compelling but describes American norms – "the way things are supposed to be" – rather than as they actually are in practice. "A far better metaphor for this track system," he explained, "is a tournament. The rule for tournament selection can be simply specified – *when you win, you win only the right to go to the next round; when you lose, you lose forever*" (Rosenbaum, 1976, emphasis in original). The model distinguished between college and non-college tracks and states that there is a continual process of selection and elimination. Each year a portion of the students are eliminated from the college tracks, and may not re-enter them in subsequent years. Although this model was certainly applicable to Rosenbaum's field site, other researchers have challenged its generalizability with findings of greater upward mobility in other contexts than the model predicts (Hallinan, 1996; Lucas, 1999; Lucas and Good, 2001). At the primary school level, teacher-created groups change with each year, and even within school years mobility is common, with upward mobility predominating over downward (see Chapter 2).

Organizational theorists described a technical model of grouping practice. In this view, grouping is a functional adaptation of schools as organizations involved in the production of individual student learning. Educators and administrators view grouping students by ability as an efficient way to deploy the "resources" available in the classroom: students, instructional materials, and time. In *Organizations in Action*, Thompson (1967) wrote that "Under norms of

rationality, organizations facing heterogeneous task environments seek to identify homogeneous segments and establish structural units to deal with each." As formal organizations tasked with maximizing student achievement with limited resources, schools face inexorable demands for efficiency. Both teachers and students report that teaching homogenous groups is easier than teaching heterogeneous ones, although few contend that teaching heterogeneous groups is impossible (Ireson and Hallam, 2001; Yonezawa and Jones, 2006). In *How Schools Work*, Barr and Dreeben (1983) explained "we can think of a district-wide distribution of children's aptitudes as a pool of resources that gets allocated as children are assigned to schools, to classes, and to instructional groups." The exact nature of the groups formed is determined by the characteristics of the student body, such as the number of students in a class or school, the mean level of student aptitude, and the amount of heterogeneity in student aptitude. Educators consider these factors and make decisions on the number of groups, their relative size, and the amount of within-group heterogeneity they desire. This will vary according to the instructional task; for instance, in elementary schools, dividing the class into several homogenous groups is standard practice for reading, but the whole class is usually treated simultaneously for math instruction.

But while Barr and Dreeben believed that "the central task of schools is to provide appropriate instruction to a large and diverse clientele in aggregations of workable size and composition" (Dreeben and Barr, 1988), other organizational theorists have pointed out that it is not the only task of schools. DeLany (1991), drawing on an institutionalist perspective (Meyer and Rowan, 1977), argued schools must also manage their environments and maintain legitimacy as well as fulfill their core instructional function. Schools attempt to match the abilities and needs of students with appropriate instructional material, but are not always able to do so given constraints on human and financial resources. Hence, students may be crowded out of particular

classes if enrollment exceeds the limits of permanent staff in a school (DeLany, 1991), be assigned to a low-level math class if they are enrolled in ESL language classes (Useem, 1992), or be shifted to or from a special education program to manage a school's standardized test score reports. The same student that is assigned or permitted to enroll in an Algebra class in one school might be allowed into an advanced Geometry in another school or relegated to a basic math class in a third. Substantial variance exists in course enrollments across different schools after controlling for student abilities and background (Garet and DeLany, 1988), and this variance is related to the social background of the surrounding community (Useem, 1992). While teachers and administrators will generally do their best to accommodate the individual needs of students, "in effect, the matches of students and courses become intelligent adaptations by the participants to the constraints placed on them. The matches are often the solutions to problems unrelated to the student's educational needs" (DeLany, 1991). Nonetheless, the strongest predictors of a student's track placement is nearly always his or her prior measures of student achievement (Kelly and Price, 2011).

Hallinan and Sørenson (1986) also proposed an organizational model described as one of vacancy competition. Individuals can advance to a higher position in a structure only when one becomes vacant. This occurs when a member of a track moves up or down, or leaves the tracking structure, and the opportunity to move into the vacancy is offered to the student in the next lower track who ranks highest in the ability distribution of that track. The authors described the consequences of this model as inequitable rather than meritocratic, because "different opportunities for learning are provided across ability group levels and assignment to an ability group level is determined not only by student ability but also by the class ability distribution and by the structural or organizational factors that affect assignment to groups." They also contrasted

their model with Barr and Dreeben's because the size of the ability groups is fixed by the teacher or school, not in response to the characteristics of their students, but by normative and structural constraints (Hallinan and Sørenson, 1986; Hallinan 1996). Unlike the institutional model of DeLany, it predicted that the same student may be assigned to different group levels depending on the particular circumstances of the class or school. In Barr and Dreeben's model, this is less likely to occur because educators have more latitude to adjust the size of groups – the organization of the classroom would adapt to the students, rather than the students' placements being adapted to fit the (locally-constrained) organization of the classroom.

Finally, and perhaps mostly importantly, some social theorists take the view that tracking and ability grouping are structures designed to confer advantage on children of affluent parents at the expense of those less fortunate. They dispute the meritocratic nature of achievement tests and teacher recommendations that are the basis for group assignment, and note that group placement is always strongly correlated with the social class background of students – poor and minority students are nearly always disproportionately assigned to lower groups. Marxist theorists Samuel Bowles and Herbert Gintis argue that the organizational structure and culture of schooling is designed to produce particular types of laborers for a capitalist economy. Students in low groups experience classroom environments that emphasize memorization and drilling, obedience to authority, and technical training in preparation to enter the work force as laborers, while those in high groups are encouraged to think critically, develop initiative, be creative, and prepare to assume leadership roles as managers and professionals (Bowles and Gintis, 1976). Empirical support for this perspective was provided by Rist (1970, 1973), who documented students being assigned to reading groups that corresponded closely to the social class origins of their families as early as the eighth day of kindergarten, and that these placements strongly

affected subsequent learning opportunities and group placements in later grades. This perspective was advanced further following the publication of Jeannie Oakes' influential monograph *Keeping Track: How Schools Structure Inequality*. Oakes (2005) links the origins of homogenous ability grouping policies in the United States with a history of influence from racism, social Darwinism, immigrant assimilation efforts, and Taylorist scientific management. She and other critics of tracking draw on a deep literature that finds that grouping benefits high-group student and is detrimental to low-group students, widening achievement inequality (Eder 1981; Oakes, 2005; Lee and Bryk, 1988; Gamoran and Mare, 1989; Hoffer, 1992; Kerkhoff, 1993; Hanusek and Wößmann, 2006; Condron, 2008). They also note that whenever detracking reforms are proposed, they are frequently opposed by affluent parents and the parents of students already identified as gifted (Oakes, 2005; Loveless, 1999; Gamoran and Weinstein, 1998). It is this perspective that has inspired calls for detracking reforms on the grounds of social justice in state and local school boards throughout the United States that continue to the present day.

Mean Effects on Student Achievement

The first research to address the issue of ability groups by directly manipulating classes was in 1916 when psychologist Guy Whipple separated groups of gifted fifth and sixth grade students into a special class, while leaving their peers in traditional classes as a control group (Whipple, 1919). In the years following, a number of similar studies were conducted. In a typical experimental study, the researcher divided students in a single grade into experimental groups composed of students of similar ability and control groups that are composed of conventional mixed-ability classes. These were accompanied by a larger number of observational studies, which compare homogenous and heterogeneous ability groups as they occur naturally in

schools systems, while attempting to control for differences between the two groups with analysis of covariance. While it would be impractical to review all of these papers here, several large meta-analyses (Slavin, 1987, 1990; Kulik and Kulik, 1992; Lou et al. 1996, 2000) have been performed on this field of research. The most widely-cited of these considers five separate types of ability grouping - multilevel classes, cross-grade ability grouping such as the Joplin plan (an early precursor to Success for All), and programs for gifted students with and without accelerated curricula. James and Chen-Lin Kulik (1992) found that over nearly a hundred studies of grouping practices, mean effects on average achievement tended to be positive and small, but statistically significant. However, there was considerable heterogeneity among the results; most studies reported positive mean effects on student achievement outcomes, a smaller number reported negative effects, and a few reported null findings.

Robert Slavin (1990) conducted a separate meta-analysis limited to ability grouping in secondary schools. The research he examined most closely resembles the multi-level classes described by Kulik and Kulik. He found that over six randomized experiments, nine matched experiments, and fourteen correlational studies, the overall mean effect size of ability grouping (versus ungrouped instruction) on achievement (as defined by the original studies) was essentially zero, consistent with the findings of Kulik and Kulik. However, Slavin excluded studies in which grouping practices were varied along with other factors, such as curriculum, teaching materials, instructional pace, or teacher training – factors that could be essential for ability grouping to be effective. Lou et al. (1996) cast a broader net and included these studies in his analysis, as well as studies of grouping at the primary, secondary, and post-secondary level. They also point out that the practice of within-class ability grouping is actually the product of two pedagogical decisions: first, whether or not to break students into smaller groups, and then

whether or not to sort students according to ability. Therefore, they consider studies that compare classes that use within-class grouping to those that use *no* grouping, and separately analyze studies that compare classes use within-class ability grouping to those that use within-class *heterogeneous* groups. For the first analysis, Lou et al. found a mean sample-weighted effect size of .17 on student achievement measures favoring classes that broke into small groups, based on 103 effect sizes in 51 studies. However, there was considerable heterogeneity among the findings. Examination of the individual findings revealed 74 effect sizes above zero, five equal to zero, and 24 effect sizes below zero. They then considered twenty findings from twelve studies that directly compared the effects of homogeneous ability grouping with heterogeneous ability grouping. The weighted mean effect size for group ability composition was +0.12, which was significantly different from zero and indicated a slight advantage to forming homogenous ability groups. This indicates that, on average, students assigned to grouped classrooms achieved more than similar students assigned to ungrouped classrooms.

Taken together, these studies provide modest support for at least some types of ability grouping. However, each set of meta-analytic findings is limited by the quality of the underlying studies. Critics of meta-analysis caution that the inclusion of substandard research undermines the validity of aggregate findings. Furthermore, the meta-analytic researchers assign equal weight to each study that meets their respective criteria. Thus, randomized experiments are treated the same as observational studies, and all observational studies are given equal weight regardless of the quality of their controls (except for Lou et al., who weighted their results by sample size). Many of the included studies have small sample sizes (less than one hundred students), short durations (as little as one semester), and many were confined to a single school (particularly among the experimental studies). The advantage of combining disparate studies is

that, with an adequate sample of studies to draw from, it becomes possible to examine which types of studies are likely to yield similar results. Hence, Kulik and Kulik distinguish between different types of grouping schemes, and are able to demonstrate the importance of curricular differentiation to finding substantial effects on learning growth. Similarly, Lou et al. contrast the importance of considering whether ability-grouped students should be compared to heterogeneously-grouped or ungrouped peers, and partially explain the variance in findings by noting factors that are associated with larger effect sizes vis-à-vis student achievement, including the use of teacher-made tests as outcomes, math and reading scores as outcomes, differentiation of materials for students, and differentiation of training for teachers.

While these meta-analytic studies provide helpful summaries of a large cross body of research, they are unfortunately limited by the quality of the underlying work on which they are based. Aggregating large numbers of findings cannot overcome the fact a large portion of research was based on correlational methods that cannot yield causal estimates, and the smaller number of randomized experiments generally had small sample sizes and short durations, limiting their generalizability. Furthermore, as Gamoron (1987) points out, neither these meta-analytic summaries nor their underlying studies provide any information about *how* grouping practices affect student achievement. In particular, they lack information regarding differences in instruction that typically coincide with differences in grouping practices. In addition to variance in methodological quality, these summaries may lump together studies that examine substantively different practices that are similar only in the purposeful creation of homogenous or heterogeneous ability groups. Nonetheless, they serve as a useful starting point when considering the widespread scope of empirical research in this field.

In the years since this earlier work was carried out, two factors have combined to substantially improve the quality of research on ability grouping. First, a number of large-scale, nationally-representative datasets, such as those gathered for the National Education Longitudinal Studies program by the National Center for Educational Statistics, have become available to the research community. Second, methodological innovations have allowed social scientists to make stronger arguments for causal effects based on observational data. One set of these, propensity score-based adjustment methods (such as stratification, matching, or inverse probability of treatment weighting) have proven popular among education researchers as a way to address the non-random selection issues endemic to education. With a rich set of covariates, a researcher can estimate the propensity for each student (or class or school, depending on the level of analysis) to experience a particular treatment. By balancing treated and untreated units on this score, it becomes possible to calculate a causal effect of the treatment under certain assumptions (Rosenbaum and Rubin, 1983). This is critical in considering the effects of ability grouping, as high- and low-ranking students usually differ so greatly to begin with that inferences about how well students placed in one group would fare if alternatively placed simply have no empirical basis. In real educational situations, high-achieving students are rarely placed in low-ranked groups (Slavin, 1990; Hoffer, 1992). Analysts must take care to ensure that common support exists between treatment groups in order to make a valid counterfactual claim. The advances in propensity score-based techniques, as well as the advent of large, nationally-representative educational datasets has led to a new generation of studies in the last few decades.

Hoffer (1992) applied traditional ANCOVA analysis to compare the achievement of students in grouped and ungrouped schools in the Longitudinal Study of American Youth, and used propensity score stratification as a specification check to validate his results. Following

students from seventh through ninth grades, he found that differences in mean math and science achievement between grouped and ungrouped schools was negligible once prior achievement and student background characteristics are controlled (although the student-level results varied depending on the group level to which students were assigned and will be discussed in more detail in the next section). In a similar longitudinal study, Mulkey et al. (2005) apply propensity score matching to adjust for treatment selection in three waves of data from the National Educational Longitudinal Study, 1998 Cohort. They found that attending a tracked middle school in the eighth grade was associated with higher mathematics test scores and substantially higher math grades in the tenth and twelfth grades. The results are independent of gender and the propensity to be assigned to a high or low track. However, Betts and Shkolnik (2000) caution that the measurement of grouping policies by survey questions in datasets like the LSAY and NELS may be unreliable if schools report that they are ungrouped because they lack a formal grouping policy, but in practice students are grouped informally or self-select into different classes.

Hong and Hong (2009) employ a similar method of compensating for selection bias, marginal mean weighting through stratification, to data from the ECLS-K. Like propensity score matching, this technique also uses a large number of covariates to weight data in observational studies to approximate what treatment groups would have looked like in a randomized experiment, with the additional advantage of robustness to certain types of model misspecification error that would undermine a traditional propensity-score based matching analysis. It also allows Hong and Hong to consider two treatments concurrently – namely homogenous ability grouping and instructional time. They find that with ample time spent on literacy instruction, ability grouping accelerated learning growth in reading, but these benefits

disappeared when instructional time was limited. Their sensitivity analysis also suggests that had they not considered the interactional effects of instructional time, they would have found no difference between homogenous grouping and whole-class instruction. Nomi (2010) also examined ability grouping in early grades using the ECLS-K data, but instead used propensity score stratification at the *school* level to adjust for the likelihood that grouping is used in different types of schools. Analyzing grouping practices at the school level has interesting property of allowing Nomi to strengthen support for part of the stable unit treatment value assumption (SUTVA) that typically plagues research on ability grouping – that the treatment assignment of one student is not affected by the treatment assignment of others. The trade off is that the second component of SUTVA – that there is only one version of the treatment – risks violation if schools implement ability grouping in different ways, which is likely true (Hong and Hong, 2009). Nomi's findings support an argument that ability grouping is an organizational response to problems of diversity in the student body. Schools that use ability grouping are likely to have heterogeneous ability compositions. They were also public, low-performing, low socioeconomic status, and high-minority schools. In these schools, ability grouping has null effects or negative effects, particularly for low-group students. In contrast, they suggest that ability grouping may improve achievement for all students in schools with advantageous characteristics, such as those found among private schools in their sample.

Several studies utilizing data gathered abroad are notable for their application of quasi-experimental techniques for causal inference with observational data. Hanushek and Wößmann (2006) combine data from several international datasets and use a difference-in-differences approach to exploit variation in the *timing* of formal grouping across eighteen countries. Some countries, such as Germany and the Netherlands, divide students into different classes or schools

relatively early in their educational careers, while others (including the United States) tend to delay formal, institutional separations until the older grades. The effect of tracking is identified by comparing performance differences between primary and secondary schools across tracked and non-tracked systems, where each country's own primary-school outcome is used as a control for its secondary-school outcome. While their findings on inequality are clear and discussed in the next section, their findings on mean achievement are inconsistent across subjects and test outcomes: countries that implement tracking earlier in students' educational careers tend to have substantially lower reading scores, somewhat lower math scores, and somewhat higher science scores. However, at least one outcome in each subject area was found to have no statistically-significant differences (Hanushek and Wößmann, 2006). The instability of results might stem from the fact that their method of operationalizing tracking as a dichotomous variable (whether a country has a policy of dividing students into separate groups before or after age 16) belies both the complexity of the grouping schemes in very different contexts and the existence of informal grouping practices such as those common as early in the first year of schooling in the United States. Huang (2009) operationalized grouping continuously, as the proportion of variance in student academic outcomes that lies between classes, an indicator of homogeneity within classrooms (irrespective of schools). He found no relationship between the amount of classroom homogeneity and mean achievement after removing country-level fixed effects.

Finally, there exists a recent, rigorously-implemented, multi-site randomized controlled trial of between-class ability grouping for young primary school students. While a number of randomized experimental studies were conducted in the past (for examples, see Billet, 1932; Drews, 1963; Marsacuilo and McSweeney; 1972), these were generally small scale studies that involved only one or a few schools. Duflo, Dupas, and Kremer's (2008) experiment involved

over ten thousand students in 121 schools in Kenya. Schools in the sample, which all had a single large first grade class, received funds to hire an additional teacher to split that class into two sections. In sixty randomly selected schools, students were randomly assigned to sections. In the remaining sixty-one schools, students were ranked by prior achievement (measured by their first term grades), and the top and bottom halves of the classes were assigned to different sections. Attrition was moderate, but even between the two treatment groups. After 18 months, students in ability-grouped classes scored 0.14 standard deviations higher than did students in heterogeneously-grouped classes, and this effect persisted one year after the program ended. They also applied a regression discontinuity analysis to examine the effects specifically for students on the margin near the cutoff score, and found that those just above the cutoff (assigned to homogenous groups) and those just below the cutoff (assigned to heterogeneous groups) both received statistically-similar benefits (Duflo, Dupas, and Kremer, 2008), suggesting that the benefits of grouping were not restricted to those assigned to the higher groups. The results provide credible evidence for the effectiveness of grouping, at least in the context studied. While there are many important differences between the Kenyan educational system and those of post-industrial countries, the research is nonetheless valuable because it is unlikely that such a randomized trial would be possible in the United States today. The controversial nature of ability grouping policy, and a greater amount of political and parental oversight, would make such large scale randomization extremely difficult to maintain in practice even if enough schools could be enlisted to participate.

Variance in Effects on Achievement by Ability Level

A primary criticism of ability grouping is that it amplifies the variance in achievement between students at different levels. In theory, ability grouping should allow instruction to be customized for each group and better targeted to the strengths and weaknesses of each set of students, allowing all students to progress as quickly as possible. In practice, however, this may not be the case. Teachers of low-ranking classes frequently have less experience than those teaching high-ranking classes (Oakes, 2005; Ansalone, 2003). Critics have argued that students in low-ranking groups are often exposed to low expectations, watered-down curricula, harsh disciplinary environments, and unstimulating instructional techniques that emphasize drilling and memorization rather than critical thinking (Bowles and Gintis, 1976; Oakes, 2005; Page, 1991; Ansalone, 2001). This suggests that ability grouping may be responsible for "Matthew effects," a term coined by Merton and adopted by education researchers (Stanovich, 1989), referencing Matthew 25:29: "For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken even that which he hath" (Bible, King James version). It may be that the small and null average effect sizes on average student achievement found in so many studies are a result of opposing influences – positive effects on high-group students and negative effects on low-group students – counterbalancing each other. Variance in the effect sizes over studies may have more to do with the underlying distributions of student achievement in the different population samples than the methodological differences between the studies. It is therefore a challenge for researchers to decompose the effects of grouping practices into separate effects on different groups. Kulik and Kulik's meta-analysis (1992) of ability grouping research also aggregated studies that examined effects on average achievement separately for students in different levels. Of fifty-one studies on multilevel classrooms, thirty-six examined results

separately by ability level, and they found slight variation. The average effect size was .10 for higher aptitude groups, -.02 for the middle groups, and -.01 for the lower aptitude students. There was moderate variance around these mean effects, and most studies reported effects sizes for all groups of less than two-tenths of a standard deviation from zero. The effect for higher groups was statistically-significantly different from zero and from those of the middle and lower groups. Slavin's meta-analysis (1990) found effects close to zero for students of all levels of prior performance.

Nonetheless, there is substantial empirical support for the hypothesis that ability grouping is associated with a widening the achievement disparities between top and bottom students (Eder 1981; Oakes, 2005; Lee and Bryk, 1988; Gamoran and Mare, 1989; Hoffer, 1992; Kerkhoff, 1993; Argys, Rees, and Brewer, 1996; Smith et al., 2008; Hanusek and Wößmann, 2006; Condron, 2008). Regarding tracking in high schools, there is evidence that the tendency for Catholic schools to enroll most or all of their students into an academic curriculum is at least partially responsible for the greater educational equity found in that sector compared to public schools (Lee and Bryk, 1988; Gamoran 1992). Gamoran and Mare (1989) applied endogenous switching regression models to control for non-random selection into tracks by allowing track assignment and track outcomes to be jointly and simultaneously determined by the data. Using data from High School and Beyond, they found evidence for differential effects on math achievement to those assigned to college versus non-college tracks, but that pre-existing differences in prior achievement and student background accounted for nearly 80% of the mean achievement gap between tracks by the senior year. The remaining gap represents a difference of about one-fifth of a SD on the outcome test. While much smaller than the raw gap, it still represents more than the total expected growth in math achievement between sophomore and

senior years (Gamoran and Mare, 1989). Hoffer's regression analysis of the LSAY found that in both math and science, "students in the high groups learn somewhat more and students in the low groups learn less than comparable students in nongrouped schools. Though the positive effects are weaker than the negative effects, more students are in the higher groups than the lower ones, and the net effect of grouping turns out to be about zero." The general pattern was confirmed when re-analyzing the data across propensity score strata, but numerous effects could not be calculated due to a lack of common support. Few students assigned to the low groups had any statistically-significant chance of being assigned to a high group and vice versa. The effects of being assigned to a low group in science were estimated using only two of the five propensity score strata (Hoffer, 1992). This indicates that the findings are not applicable to all students, but restricted to those closer to the margins between track assignments, and underscore Slavin's (1990) caution in interpreting and generalizing from low-high group comparisons.

Support for the Matthew effects hypothesis in the early grades can has also been found by those examining ECLS-K data. Tach and Farkas (2006) used hierarchical linear modeling to account for the clustering of students in classes, but relied on a limited set of covariates to compensate for selection bias in group effects. Treating ability group placement as a continuous variable, they found that "ability group placement and the teacher's assessment of student behavior both have significant effects on students' growth in reading achievement, even net of their prior reading achievement scores. Such grouping takes individual- and group-level performance differences that emerge during the preschool period and causes them to widen more than would otherwise be the case during the first two years of formal schooling." Condron (2008) followed by using propensity score-based stratification analysis to separately compare the effects on students placed in high, middle, and low reading groups with students with similar

propensity scores who were taught in ungrouped classes. He found that placement in a high group yields an average benefit (across five propensity strata) of an additional one-sixth SD of reading growth in first and third grades, placement into a middle group is ambiguous across strata but never statistically significant, and placement in a low group results in a loss of about one-fifth of a SD in reading growth over first and third grades (Condron, 2008). However, Hong et al (2012) contended that the strategy of calculating propensity scores for each ability group separately is flawed because it misrepresents the selection process. Which group a student would be assigned to depends on the student's relative standing in the ability distribution of the entire class. In response, they applied MMW-S to create a weighted "pseudo-subpopulation of children at each ability level." Within each of these weighted subpopulations, the kindergarten classes attended by these children can be viewed as if they had been randomly assigned to different instructional treatments (ungrouped or ability-grouped classes), conditional on an assumption that no unobserved confounding variables exist. They also interacted this treatment with the amount of instructional time students receive in literacy. Their results suggested that, under conditions in which students are afforded ample time for literacy instruction, homogenous ability grouping did not substantially change the reading outcomes of high- or low-ability students, but that middle-ability students *benefited* from grouping with an effect size of about .12 SD. However, when instructional time was limited, the detrimental effects of ability grouping on low-group students became apparent (Hong, 2012). These results were consistent with the argument of Gamoran and others that the effects of grouping will be contingent on the instruction provided to different groups.

The evidence from the economic tradition is mixed but provides some support for differential effects. First, several recent studies support the Matthew effects hypothesis.

Hanushek and Wößmann's (2006) difference-in-differences analysis of cross-national data found that nearly every country that differentiated students by ability before sixteen years of age experienced increased academic inequality between late primary and secondary schooling. Conversely, seven out of nine countries (including the United States) that were identified as implementing ability grouping later in a child's educational career saw decreases in academic inequality between primary and secondary schooling. The results held regardless of how inequality was operationalized (in terms of standard deviations, 25th-75th percentile gaps, or 5th - 95th percentile gaps). Another interesting study by C. Kirabo Jackson (2009) exploited a rule-based student assignment mechanism in Trinidad and Tobago. Jackson used the complex algorithm designed to assign students to secondary schools as an instrumental variable in order to isolate exogenous variation in the schools that students attend. The policy also created sharp cut off points that allowed the use of regression discontinuity analysis to compare the outcomes of very similar students that are assigned to different treatments (in this case schools with different average peer abilities) by virtue of a strict policy. The findings from both techniques were convergent: on the margin, students did better when grouped in a school with higher-ability peers than they will in a school with lower ability peers. This suggests that "ability grouping reinforces achievement differences by assigning the weakest students to schools that provide the least value-added" (Jackson, 2009).

Several other papers stand against the Matthew effects hypothesis. Figlio and Page (2002) used OLS regression with NELS data and find that, when comparing the coefficients of high- and low-track dummy variables in regression models (a common methodological choice, see Hoffer, 1992 as an example), mean differences are significant even when controlling for prior achievement, suggesting differential effects of ability grouping. But the authors suggested

that these results are not causal, due to unmeasured endogeneity of track assignments. As an alternative, they stratified the sample into low-, middle-, and high-ability groups on the basis of students' 8th grade test scores rather than their high school track assignment. They found that none of these three groups experienced any detrimental effects when they attend schools that use tracking versus those that do not. They further argue that both of these analyses are flawed because they do not consider the possibility of endogeneity of the *school choices* in track assignments. They used a set of complex instrumental variables that represent factors that would influence a parent's decision to place their child in a tracked or untracked school: policy environment, school environments, and parental tastes. Their chosen instruments were unfortunately weak, so the resulting causal estimates are imprecise, but they suggest that "the effect of tracking is, if anything, *positive* for members of the low-ability group." The results were robust to several different specifications of tracking. Additional support for the idea that tracking benefits low-ability students as well as high-ability students comes from Duflo, Dupas, and Kremer's (2008) randomized controlled trial in Kenya. In schools that implemented ability grouping, students in the lower group benefited just as much as those in the higher group. The effect size was about one-sixth of a SD positive difference in both math and reading scores, compared to students in ungrouped classes. Furthermore, by disaggregating the test scores and examining changes in individual competencies, they found evidence that suggest the benefits were generated by giving teachers the opportunity to customize instruction to focus on specific student needs. While there was no clear pattern in language arts achievement scores, they found that children assigned to lower groups gained from tracking more than other students on the easiest questions and less on the more difficult questions. Conversely, students assigned to the higher group did not significantly benefit from tracking for the easiest questions, but they did

significantly benefit from it for the hardest questions (Duflo, Dupas, and Kremer, 2008). It may be that ability grouping provides benefits to all students when implemented in an ideal way, but that the social and organizational consequences of tracking students in typical educational systems (discussed in the section on "Proposed Mechanisms" below) undermine these benefits.

Grouping and Learning Behaviors

One limitation of the existing literature on grouping is that very few studies consider student learning behaviors either as a factor contributing to group assignment or an outcome influenced by students' experiences in grouped instruction. Most extant studies do not have information on these behaviors, and rely primarily on prior academic outcomes and sociodemographic controls when attempting to identify the effects of group assignment on academic outcomes. Consider the following proposition: Through their daily interactions with students, teachers know which students display better learning behaviors, while researchers do not. Teachers use this information to help make group assignments, assigning students with better learning behaviors to higher groups. If the students themselves have input in the decision making process (say, when offered a choice between regular and advanced courses in high school), they also use information on their work habits when making decisions. If, as we might expect, students with better learning behaviors are sorted into higher groups and have better subsequent academic outcomes, researchers will over-estimate the effects of group assignments unless differences in learning behaviors are explained by other information available to researchers. This might (at least partially) explain the prevalence of Matthew effects in the literature – students with better learning behaviors and higher learning growth rates have unobserved selection into higher groups.

There are good reasons to believe that learning behaviors are important factors that influence group assignments even conditional on prior student achievement. Ethnographic studies of kindergarten classes have documented students being assigned to reading groups before they are ever formally assessed by exams, partially on the basis of behavioral cues, most notably their ability to sit still and pay attention (Rist, 1970; Eder, 1981). Haller and Waterman (1985) conducted interviews with sixty primary school teachers and asked them to form recommended reading groups for the following year. They found that "even though all were creating ability groups, not one respondent was solely concerned with ability." Reports of students work habits also influenced teacher's group recommendations, especially in marginal cases where a the teacher's perception of the child's abilities suggested placement in either a higher or lower group would be plausible. They explain, "Concerning work habits, a frequent rationale was that a child would (or would not) work hard enough to keep up with the group being considered. In regard to behavior/personality, motivation was cited most often as the reason for the decision." At the high school level, Finley (1984) found an even greater importance placed on student effort:

> Teachers of advanced classes stress two student characteristics, ability and motivation, as essential in these good classes, but students are distinguished primarily by motivation. Teachers have discretion in allocating students to tracks, and students are sorted into classes more by motivation than by ability... Many teachers mentioned that they have a few "bright kids" in lower-track classes, but these students do not want to do the work, so they seldom recommend that such students transfer to higher-track classes. On the other hand, teachers mentioned that there are always a few "borderline kids" in advanced classes who are not really advanced but who are allowed, even encouraged, to remain in the high track because they are motivated and fit in.

Carbonaro (2005) found large between-track differences in teacher reports of student effort, and that these differences in turn explained a portion of the between-track gaps in student achievement. Similarly, Smith (1998) found that students' effort in math was correlated with ability group placements in middle schools.

Another possibility is that the different learning environments encountered by students in different groups shape the development of learning behaviors. This could mean that improvements in achievement outcomes associated with group assignment are mediated by changes in learning behaviors. Learning behaviors might be influenced by a number of factors associated with group assignments, discussed in more detail in the next section. Empirical research evaluating the proposition that grouping practices affect learning behaviors is relatively sparse compared to research on academic outcomes. Focusing on kindergarten and first grade, Tach and Farkas (2006) found support for the Matthew effects hypothesis, showing that assignment to a higher group was associated with better learning behaviors after controlling for prior behaviors and other covariates. Catsambis and Buttaro (2012) extended this work with propensity-score based matching analysis. Their results indicated that students assigned to higher groups experienced greater positive changes in several aspects of pyscho-social development in kindergarten, including learning behaviors, compared with matched non-grouped students, while students assigned to low groups developed less. However, findings from Hong et al. (2012) focusing on the same students but using a marginal mean weighting strategy, presented a more complex picture. Hong et al. found that the impact of grouping on learning behaviors depended both on the ability levels of students and the amount of time dedicated to grouped reading instruction. For middle-ability students, high-intensity grouping improved general learning behaviors when it was combined with ample instructional time, but not when time was limited. For low-ability students, grouping was undesirable in both contexts, but no significant effects were found for high-ability students. Nomi (2010) took a different approach by comparing grouping practices at the school-level. Descriptive analyses revealed that that while schools where all students were in grouped classes had similar average learning behaviors as

schools in which no classes used grouping, the later had significantly less variance in student learning behaviors.

Proposed Mechanisms

If ability grouping does have differential effects by group level on academic outcomes, there are several proposed mechanisms through which they may operate. While there is little existing literature examining how group assignments affect learning behaviors, we can review research that documents differences in instructional environments between groups and studies that address differential effects on achievement. I do so here, emphasizing findings that might be extended to suggest ways that group assignments can change students' learning behaviors.

The most frequently-suggested factor is that the curricular coverage of lower-ranking groups is both quantitatively and qualitatively inferior to that of the higher groups. Students in low groups cover less content over the course of an academic year, denying students opportunities to learn (Oakes, 2005; Page, 1991). A student cannot learn material to which he or she is not exposed. Students in the low-ranking groups experience more interruptions and spend less time on task than their peers in other groups, both in within-class groups (Allington, 1983; Eder, 1983; Chorzempa and Grapham, 2006; Grant and Rothenberg, 1986) and tracked classes (Schwartz, 1981; Finley, 1984; Page 1991). Perhaps most importantly, instruction in low-ranking groups tends to emphasize memorization and simple problem solving, whereas instruction in high-ranking groups is more likely to include an emphasis on critical thinking skills and complex analysis (Bowles and Gintis, 1976; Oakes, 2005; Page 1991). This is true even when the same teacher instructs classes of different levels (Raudenbush, Rowan, and Cheong, 1993). Students are more likely to engage with lessons they perceive as valuable and

challenging (although not impossible) and less likely to engage with tasks they perceive as simplistic or tedious (Stipek, 2002).

It is possible the dearth of higher-order instruction in low-ranking groups is a result of lowered expectations. While interviewing reading teachers in a comprehensive high school, Finley (1984) reports, "When they first had remedial classes, they tried their favorite methods of large-group discussion and small-group projects (used extensively with high-track classes) but found that students did not respond. They gave up such methods early on and devised lesson plans based upon individual worksheets for drill in English mechanics." Unfortunately, instructional differences have substantial influences on student outcomes. Carbanaro and Gamoran (2002) found that the study of literature and analytic writing improved measured learning growth in reading achievement, while lessons on grammar actually *decreased* growth rates. Instruction-by-track interaction terms were small, suggesting that the more basic techniques that characterize low-track instruction were not actually more effective, but rather limited the learning opportunities of students.

It is also generally theorized that students in low groups or tracks suffer from a stigma attached to a low-status group in an academic context. While this might not be a serious problem for very young students (nearly all kindergarten-age students perceive their own academic abilities as very high), it becomes an increasing concern as students develop a more accurate understanding of their own abilities, their standing relative to their peers, and the meaning of group distinctions as they age through the primary grades (Dweck, 2001; Eder, 1981; Stipek, 2002). In a social network survey of 3rd and 4th grade students, 50 percent of low-rank students chose more high-rank than low-rank students to "hang out with the most," but less than one percent of high-rank students chose more low-ranking students. Students in low groups had less

than half the number of reciprocal friendship choices than their high-ranked peers (Schwartz, 1981). Student friendship networks also tend to align with ability groups over time, increasing social distance between students in different groups (Hallinan and Sørenson, 1985). Moreover, the social status attached to students on the basis of group assignment may extend to their instructors in schools that use between-class grouping. Teachers report valuing students in high-track classes for their compliant behavior, high levels of motivation, and the fact that their parents are often educated and have high-status occupations. The teaching of low-track students is also sometimes stigmatized by other teachers, and being repeatedly assigned large numbers of low-track students can be a source of resentment. Conversely, teaching a high-track class is considered a high-status position, and is often viewed by those that hold it as a meritocratic reward for professional competency (Pink and Sweeney, 1978; Finley, 1984; Ireson and Hallan, 2001).

Assignment to low-ability groups may also cause students to internalize negative assessments of their ability, resulting in discouragement. Surveys suggest that students in low tracks have substantially more negative views of themselves both academically and generally, and have substantially lower educational expectations than those in higher tracks (Oakes, 2005). In observations of primary school students, students in middle- and high-ranking reading groups were allowed to jump in and finish sentences for other students, while students in lower-ranked reading groups were reprimanded for doing so (Eder, 1981). Pink and Sweeney (1978) report that in interviews with students in the 7th grade, "Almost without exception, students in the [low] track volunteered the information that they were either 'dumb' or 'ain't smart.'" Lee and Bryk (1988) found that among students who indicated in 8th grade that they plan to attend college, only about one half were in an academic track by their sophomore year and only about a quarter

remained by senior year. Even in schools that implemented efforts to detrack by removing prerequisites requirements and limits on course enrollment, few students in lower-ranked classes took advantage of the opportunity due to uncertainty over how they would handle the greater challenges. Others were discouraged by school personnel or peers from taking classes perceived to be beyond their abilities. This effect was strongest for black and Latino students, who worried that their contributions would not be valued in classes dominated by white, upper-class peers (Yonezawa, Wells, and Serna, 2002). Gouldner (1978) documented classrooms where students assigned to the lowest group were seated farthest from the teacher. As the school year progressed, the teacher responded less to these students, and they in turn responded less to the teacher.

On the other hand, the stigma and discouragement associated with being assigned to a low group and the honor of being placed in a high group are not the only ways a student's self-concepts (i.e. perceptions of their own academic abilities) may be affected by the placement. Students will also contrast themselves to others *within* their group. A low-ability student in a heterogeneous group or class may not only struggle with material presented by instructors "teaching to the middle" but also face high-ability classmates that assimilate the material more easily. Conversely, a high-achieving student may compare their abilities favorably to the majority of students in a heterogeneous class, but be discouraged by comparisons of their abilities to those of their highly-capable peers in a high-track class. Marsh (1987) terms this the "Big Fish Little Pond" effect in the psychological tradition, and Richer (1976) considers it an effect of defining a student's "reference group" in the sociological tradition of Merton. It is possible that these contrast effects would have a stronger influence on students' academic self-concepts than the stigma or glory of group assignments, because interactions within the

immediate social circle should figure more prominently into self-conceptions than those of more socially-distant groups. Chiu et al. (2008) found that students in higher tracks had greater confidence in their math skills and academic abilities in general, although not after controlling for grades. Furthermore, math track did not influence students' self-esteem. Students reported that they most often compare themselves to those of similar ability in the same track, and those that did compare themselves to other students in higher or lower tracks did not show substantive effects on their self-esteem.

Mulkey et al. (2005) found that attending a tracked middle school had significant negative effects on the academic self-concepts of high-achieving students (that is, students with a high propensity to be in high-track class). High-performing students in tracked schools in 8th grade had lower educational expectations in 10th grade than those that attended untracked schools. Conversely, students with a low propensity to be placed in a high track had higher educational expectations in 10th grade if they attended a tracked middle school, controlling for student covariates and propensity scores. These effects were considerably stronger in boys than girls (Mulkey et al, 2005). Even in Germany, where students are grouped by ability in a highly-visible way into a tiered system of secondary schools, contrast effects had a much greater influence on students' academic self-concept and interest than the effects of track assignment, once individual student achievement, group-average achievement, and teacher-assigned grades were controlled (Trautwein et al., 2006). Kulik and Kulik's meta-analysis (1992) documented eleven studies of multilevel classes that separated results on self-concept and self-esteem by ability level and found a similar pattern: the average effect size is -0.15 on high-aptitude students, -0.09 on middle-aptitude students, and 0.19 on low-aptitude students. However, it may be the case that within-class ability grouping, as is common in primary schools in the United

States, facilitates easier comparison of oneself to peers in different groups, exacerbating the stigma associated with being assigned to a low group and lowering expectations of students assigned to them (Reuman, 1989), although other researchers have found null effects of within-class grouping on academic self-concept (Pallas et al., 1994, Lou et al., 1996). Mulkey et al. (2005) also note that "middle school tracking's links to high school test scores are independent of student gender and propensity for high or low track, therein suggesting that the subjective social and institutional aspects of tracking are less influential in determining test proficiency than are the instructional aspects of tracking." The subjective evaluation of students' abilities *by teachers* may also be influenced by within-class comparisons. When algebra reforms in Chicago substantially increased within-class heterogeneity of students' prior math achievement, the grades of students in the lowest part of the distribution suffered and course failure rates increased (Allensworth et al, 2009).

Finally, it has been suggested that the effects of ability grouping on achievement may operate through exposure of students to different types of peers. The students assigned to higher-ranking groups tend not only to display a greater aptitude for the subject matter but also more motivation, better behavior, and greater enjoyment of the material than those assigned to lower groups. Behavioral norms in high-ranking groups are generally more aligned with the expectations of teachers and more conducive to learning (Eder, 1981; Oakes, 2005; Page, 1991). Duflo, Dupas, and Kremer's (2008) randomized experiment in Kenya also considered peer effects by examining random variation in mean peer ability in heterogeneous classrooms. By regressing student outcomes on the mean pre-test outcomes of their peers in each class, the economists estimate a standardized effect size of about .53. The total effect is a combination of a direct effects of peer ability and an indirect effect through changes in teacher behavior.

Therefore, while interesting, estimates of the effect of local variation in heterogeneous classes do not capture endogenous teacher responses to the changes in class composition generated by tracking. Regression discontinuity results indicated students just below the cutoff score were not harmed by grouping, suggesting that the indirect benefits of proper instruction can overcome the negative effects of lower-ability peers. In contrast, Zimmer (2003) analyzed observational data from the United States and found that, in practice, peer effects are dominant for students in low-ranking tracks. Peer ability and ability grouping interact non-linearly, but exposure to high-ability peers is clearly beneficial. He notes that while his regression models indicate that tracking does have a net positive affect for low- and middle-rank ability students at most levels of peer achievement, the question of whether students will perform better in tracked or untracked classes will depend on the average ability level of their peers. His results are consistent with the Matthew effects hypothesis as they "suggest that the use of tracking diminishes the impact peers have on student achievement for low- and average-ability students while the peer effect is unaffected by tracking for high-ability students" (Zimmer, 2003).

Many authors in the literature have remarked on how behavioral norms and the tone of interaction can differ substantially between achievement groups, and the environments of the lowest groups are particularly disruptive of learning (Eder, 1983; Gouldner, 1978; Oakes, 2005; Page, 1991). Although it is possible that these differences might be due to pre-existing behavior patterns of students sorted into different groups, there are reasons to believe that groups develop emergent norms that can encourage or discourage learning-related behaviors. An illustrative example comes from observations of a first grade classroom, where Felmlee and Eder (1984) documented declining attention to lessons in lower groups as the year progressed, even after controlling for student background characteristics, prior attention, reading ability, and teacher

management. One student ("Zach") was reassigned from the medium-low group to the medium-high group during the year. When he was in the medium-low group, he was more often distracted by behavior of his peers, and the teacher had to spend more time correcting the behavior of his peers than addressing Zach's inattentiveness. After Zach was moved, students in the new group socialized him to new group norms, commenting on inattentive behavior and reminding him to read in his head versus aloud. Moreover, "the teacher's response to Zach's inattentive behavior was dramatically different from her response to his behavior when he was a member of the medium-low group. This no doubt is due in part to the fact that because there was much less inattention in general in the medium-high group, the teacher can respond more quickly to inattentive acts when they do occur without having to be continuously managing students' behavior." The combination of changes in teacher expectations and peer influences in the new group improved Zach's attention to lessons substantially, reducing his inattentive behavior from 43% of the time before the move to only 18% afterwards. At the high school level, students' experiences in tracked instruction can result in polarization in student academic engagement and behaviors (Berends, 1995; Rosenbaum, 1978; Schwartz, 1981). My task is to investigate whether this potential divergence in learning behaviors prevents upward mobility for students assigned to low-ranking groups early in their careers, or whether determined students can work their way up as they progress through school.

Chapter 2 – Within-class grouping in the early primary grades

Introduction

Within-class grouping is the first ability grouping practice encountered by most students in the early grades[1]. While Rist (1970) and Gouldner (1978) observed students divided into permanent, physically-separated groups for lessons in all subjects, students are typically grouped for only a portion of the school day, most often for reading instruction, but sometimes also for math or other subjects (Barr and Dreeben, 1983; Eder, 1983; Chorzempa and Graham, 2006; Schumm, Moody, and Vaughn, 2000; Weinstein, 1976). Nonetheless, there are substantial concerns that, once students are assigned to low-ranking groups, lowered expectations and inferior instruction make it more difficult for students to develop as readers. As reading is a fundamental skill for learning other subjects, this can have a profound impact on students' development through the elementary grades. Furthermore, because black and Hispanic students and students from poor families are disproportionately assigned to low groups, critics argue that the practice places these students at a further disadvantage nearly as soon as they enter school (Allington, 1983; Catsambis and Buttaro, 2012; Rist, 1970, 1973; Schwartz, 1981; Stanovich, 1976; Strike, 1983; Tach and Farkas, 2006).

In this chapter, I consider the iterative process where students are assigned to reading groups (on the basis of prior achievement, learning behaviors, and other factors), group assignments shape the development of achievement and learning behaviors, which in turn shape subsequent group assignments. While some earlier researchers have found that, once students are assigned to low groups, it is nearly impossible to move up (Austin and Morrison, 1963; Gouldner, 1978; Rist, 1970, 1973), subsequent work has generally found that mobility between reading groups during a single school year is possible if somewhat limited (Clark-Ibanez, 2005;

[1] That is, unless one considers the formation of age-graded classrooms to be a type of ability grouping.

Dreeben, 1984; Hallinan and Sorensen, 1983; Eder, 1983; Weinstein 1976). Because students move to new classrooms with new teachers each academic year, it is reasonable to suggest that these transitions might provide even greater opportunities for students to be reclassified and move either up or down. Research on this part of the process is extremely limited. Rist and Gouldner followed students from kindergarten through second grade, and found almost no mobility as students progressed. Using data from six schools in Chicago, Gamoran (1989) considered how student achievement and first grade reading group ranks predicted second grade reading group ranks. He found that only first grade reading group assignments predicted second grade assignments in three schools. In two schools only student achievement predicted second grade assignments, and in the final school, both did. The analyses presented in this chapter represent the most systematic investigation of group mobility in the early primary grades to date.

I also extend existing work describing how assignment to within-class reading groups affects the development of student learning behaviors. Qualitative researchers have observed classrooms and documented changes in behavior patterns over time among students assigned to different groups. They suggest that differential instruction, expectations, and peer effects combine to create self-fulfilling prophecies where positive learning behaviors are reinforced for members of high groups and undermined for students in low ones (Allington, 1983; Eder, 1981; Eder and Felmlee, 1984; Gouldner, 1978). Others have used large-scale longitudinal datasets, including the ECLS-K dataset analyzed in this chapter, to compare the learning behaviors of students in grouped and ungrouped classrooms. Tach and Farkas (2006) found that evidence of Matthew effects on learning behaviors for students assigned to different groups after adjusting for covariates. Catsambis and Buttaro (2012) corroborated these findings using propensity-score based matching. Hong et al. (2012) found that the impact of grouping on learning behaviors

depended both on the ability levels of students and the amount of time dedicated to grouped reading instruction. For middle-ability students, high-intensity grouping improved general learning behaviors when it was combined with ample instructional time, but not when time was limited. For low-ability students, learning behaviors showed greater improvement with more instructional time, but no significant effects were found for high-ability students. However, each of these studies focused on comparing students assigned to different groups with counterparts in ungrouped classes. The work below has a different counterfactual focus: given that students encounter grouped instruction in their classrooms, how does assignment to a higher or lower group affect the development of academic achievement and learning behaviors?

Specifically, I address three research questions:

1. What student background, academic, and learning behaviors predict students' group assignments?
2. How much mobility between group levels occurs, and what academic and behavioral characteristics predict this mobility?
3. How does assignment to a high or low ability group affect academic outcomes and subsequent learning behaviors?

Data

Data come from the Early Childhood Longitudinal Study - Kindergarten Cohort of 1998 (ECLS-K). Initiated by the National Center for Education Statistics, the ECLS-K collected information from a nationally representative sample of kindergartners, their parents, teachers, and schools across the United States. It followed the same children from kindergarten through the 8th grade. Information was collected in the fall and the spring of kindergarten (1998-99), the

fall and spring of 1st grade (1999-2000), the spring of 3rd grade (2002), the spring of 5th grade (2004), and the spring of 8th grade (2007). Children, their families, their teachers, and their schools provided information on children's cognitive, social, and emotional development. Information on children's home environment, home educational activities, school environment, classroom environment, classroom curriculum, and teacher qualifications was also collected. The analytic sample includes 13,380 students in 3,899 classrooms in 1,339 schools in the first grade wave of data collection. Attrition reduced the number of students to 11,110 students in 4,615 classrooms in 2,140 schools in 3rd grade. (Note: All sample sizes throughout this dissertation have been rounded to the nearest 10 to protect respondent confidentiality.)

Data on grouping assignments comes from teacher survey questions concerning sampled students. Teachers reported whether or not they used ability grouping for reading, how many groups they used, and the group assignment of sampled students. From this information, I operationalize group placement three different ways: a dummy variable for highest group membership (coded 1 if the student is a member of the highest group, 0 if the student is in any other group), a dummy variable for lowest group membership, and a continuous measure of group placement proposed by Pallas, Entwisle, Alexander, and Stluka (1994). The continuous measure creates a "decile-like" summary measure of group assignment that is adaptive to classrooms with different numbers in the groups. It is calculated as follows:

$$\frac{10}{n_{groups}} \times (n_{lowergroups} + \frac{1}{2})$$

where n_{groups} is the number of groups in the class and $n_{lowergroups}$ is the number of groups lower than the group the target student is placed in. For example, in a classroom with two groups, students in the higher group would be assigned a value of 7.5 and students in the lower group would be assigned a value of 2.5. In a classroom with four groups, however, students would be

assigned values of 8.75 in the highest group, 6.25 in the higher middle group, 3.75 in the lower middle group, and 1.25 in the lowest group. Between-year group mobility is coded as the difference between these continuous measures. Students in ungrouped classes in the prior year were coded as having a continuous group measure of 5, and a separate indicator variable for prior ungrouped status was included in the model. In each transition period (between kindergarten and first grade, and between first and third grades) this mobility measure has a mean close to zero and a standard deviation of about 2.7. The continuous measure has particular advantages here due to changes in the number of groups students might be assigned to across classrooms. Consider two students that were assigned to the lower of two groups in kindergarten. The next year, one moves into a classroom with two groups, and the other moves into a classroom with five groups. *Ceteris paribus*, the first student will have a considerably larger chance of moving into the highest group than the second. If we consider only whether students move into the highest (or lowest) groups, then there is a mechanistic relationship between the number of groups in the classrooms and opportunities for mobility. For example, across all students who were in classrooms with two groups in kindergarten and moved into classrooms with five groups in first grade, upward mobility into the highest groups in first grade occurs for only 1.3% of students. Conversely, for students moving from a kindergarten classroom with five groups into a first grade class with two groups, the probability of upward mobility into the highest group is 43.6%. Therefore, mobility into and out of the highest and lowest groups is mechanistically related to the number of groups formed in each year.[2] However, using the continuous measure, average mobility is -.006 and -.100 (not statistically

[2] This might not be a serious concern if the numbers of groups formed in the classrooms students enter are independent of student background and prior treatment conditions. However, ancillary analyses suggested that this is not necessarily true. More advantaged students, students with better prior reading scores, and students that attended kindergarten classes with a larger number of groups were all slightly more likely to be placed in first grade classes with a larger number of groups.

different from zero) respectively. I also conduct a supplementary mobility analysis that restricts the sample to those making non-structural moves between years – that is, students that moved up or down when they attended classrooms with the same number of groups in each year.

I consider three outcome variables of primary interest. The first is the item response theory theta for the reading test administered by ECLS-K staff. This measure estimates the underlying reading ability of a child on a continuous scale by evaluating correct, incorrect, and skipped questions of various difficulties on a standardized reading test. The test was based on the 1992 and 1994 National Assessment of Education Progress Reading Frameworks (ECLS-K Psychometric Report). Second is a summary measure of learning behaviors from the teacher survey called the Approaches to Learning subscale. It is composed of items asking teachers to indicate level of agreement with the following statements to describe sampled students:

- Keep working at something until {he/she} is finished?
- Show interest in a variety of things?
- Concentrate on a task and ignore distractions?
- Help with chores?
- Eager to learn new things?
- Creative in work or in play?

Each response scale ranges from 1 to 4. The Approaches to Learning subscale was calculated only if there were valid data on at least four of the six items. The subscale is the mean of all items with data. It has a split-half reliability of .89 (ECLS-K Psychometric Report). The final outcome measure is the single 1 to 4 scale response for the teacher survey item "How often does this child work to the best of her/his ability?"

I also make use of a large number of student, class, and school-level covariates. Descriptive statistics for these variables from the first wave of data collection are included in Appendix 1. Multiple imputation was used to address missing data on covariates. Information on treatments (grouping assignments) was used in the imputation of covariates, but no treatment

variables were imputed. Among students that had data on treatments and were included in the analytic sample, missing data rates on covariates were modest to moderate (see Appendix 1). Five imputed datasets were created using the Markov Chain Monte Carlo method with PROC MI in SAS. Each wave of data was imputed separately, but information from prior waves was used in the imputation of later waves. A ridge prior with two degrees of freedom was used to alleviate multicollinearity among the covariates and allow the imputation models to converge. The addition of this prior slightly attenuates the correlation among the many variables in the model without changing the means or variances of their distributions. It is analogous to adding two cases to the data from a hypothetical population with uncorrelated variables (Enders, 2010; Schafer, 1997). At the time of this writing, proper implementation of multiple imputation is still being developed for the HLM software. I have chosen to report findings based on a single imputed dataset. Because this does not account for uncertainty in the imputed values (and attendant variation in estimates) the standard errors are slightly underestimated. The findings presented are substantively similar to estimates using all five imputed datasets modeled in a single-level framework in SAS.

Methods

My first two research questions, concerning the description of predictors of group and mobility between groups are primarily descriptive and do not require adjustment for covariates. However, the question of whether group assignment causally affects learning behaviors and academic outcomes is more difficult because treatment depends so heavily on prior outcomes. To address this, I use propensity score stratification to balance students assigned to different groups on observed covariates (Rosenbaum and Rubin, 1984). Because I am interested in the effects of

group assignment given the use of grouping in the classroom (rather than comparing group assignment versus ungrouped instruction), I limit the analytic sample for these analyses to students in classrooms that employed grouped reading instruction.

This approach approximates the results of a randomized experiment under certain assumptions, namely strong ignorability and the stable unit treatment value assumption (SUTVA). Given the large, high-quality set of covariates present in the ECLS-K, the strong ignorability assumption is plausible – that is, conditional on observed covariates, treatment assignment is independent of the potential outcomes for each student. Sensitivity analysis is used to evaluate how robust the findings are to omission of potential unobserved confounding variables. The stable unit treatment value assumption, on the other hand, is more problematic. Both parts of this assumption – that there is only one version of the treatment, and treatment assignment of one student does not affect the outcomes of others – are not likely to reflect the reality of grouping practices in education. It is unclear how violation of this assumption may bias the results of this research. Because the sample was subdivided into many strata, analyzing the data in hierarchical framework became problematic. The average classroom sample size across the analytic samples was about 3, and the average school sample size was about 7, making it difficult to subdivide these subpopulations in a way that balanced students assigned to different groups on all covariates. However, because classroom and school environment are important predictors of both group assignments and academic outcomes, variables on both were included in the propensity models. I also present a separate set of propensity score analyses that includes school fixed effects (and omitting school-level covariates) to account for potential unobserved school factors influencing student outcomes.

Results

What student background, academic, and learning behaviors predict students' group assignments?

Hierarchical regression models with key variables predicting group assignments are detailed below. Students are nested within classrooms (level 2) and schools (level 3). All covariates are centered around their grand means, and all covariate slopes are fixed (intercepts at each level vary). Logistic regression models use Penalized Quasi-Likelihood estimation,[3] and unit-specific estimates are reported. Group assignments in 1st grade are described in tables 1.1 through 1.3 and assignments in 3rd grade are described in tables 2.1 through 2.3. In each table, Model 1 includes measures of reading achievement, learning behaviors, and group assignments from the prior wave of data. Model 2 adds student-level background characteristics, and Model 3 adds classroom[4] and school context variables.

[3] EM LaPlace estimation is generally preferred for models with dichotomous outcomes in HLM, but the estimation failed in a number of models described in the tables below, possibly due to small level 2 and 3 sample sizes.

[4] Note that I have included a measure of classroom mean achievement (the percent of the class reading below grade level) in this model. This is problematic, because class mean achievement could have been effected by the treatment, grouping practices in the classroom. However, I have elected to include this variable because it provides important contextual information for students' group assignment. The inclusion of this variable does not appear to introduce much bias into the other estimates, as all other coefficients change very little when it is omitted.

Table 1.1 – Predictors of Highest Group Assignment in 1st grade

	Model 1		Model 2		Model 3	
Variable	Coeff.	Std Error	Coeffi.	Std. Error	Coeff.	Std Error
Intercept	-1.066**	.048	-1.072**	.047	-.994**	.046
Prior Reading Score	3.096**	.094	3.152**	.099	3.318**	.101
Prior Learning Bhvrs	.402**	.068	.397**	.069	.423**	.070
Prior Works Best	.240**	.067	.231**	.067	.229**	.067
Highest Group in K	1.032**	.102	1.025**	.102	.967**	.103
Lowest Group in K	-.134	.154	-.143	.154	-.134	.155
Ungrouped in K	.461**	.094	.486**	.094	.549*	.094
Male			-.058	.060	-.084	.061
Black			.270*	.111	-.169	.133
Hispanic			-.001	.106	-.350*	.117
Asian			.143	.149	.124	.150
Other Race			.193	.169	.111	.172
Non-English Home			.340**	.120	.053	.126
Special Needs			-.047	.113	-.050	.114
Siblings			-.023	.027	-.044	.028
Disabled			.011	.090	-.008	.092
Single Parent			-.066	.083	-.105	.084
SES			-.077*	.037	.021	.039
Class Size					-.048**	.010
Class % Male					.007*	.004
Class % Black					.010**	.002
Class % Hispanic					.015**	.002
Class % Reading Below Grade Level					.012**	.003
Private School					-.994	.047
School Reading %ile					.120	.154
School Math %ile					.013*	.007
School % FR lunch					-.017*	.007

Table 1.2 – Predictors of Lowest Group Assignment in 1st grade

	Model 1		Model 2		Model 3	
Variable	Coeff.	Std Error	Coeffi.	Std. Error	Coeff.	Std Error
Intercept	-1.609**	.045	-1.614**	.045	-1.712	.042
Prior Reading Score	-2.393**	.091	-2.393**	.097	-2.582**	.101
Prior Learning Bhvrs	-.540**	.066	-.498**	.067	-.518**	.068
Prior Works Best	-.161*	.069	-.166*	.070	-.186**	.070
Highest Group in K	-.918**	.152	-.923**	.152	-.837**	.153
Lowest Group in K	.445**	.112	.449**	.113	.414**	.114
Ungrouped in K	.123	.090	.092	.090	.028	.091
Male			.006	.064	.016	.065
Black			-.250*	.108	.005	.134
Hispanic			-.163	.110	.197	.122
Asian			-.191	.175	-.175	.177
Other Race			-.218	.184	-.148	.185
Non-English Home			-.395**	.125	-.105	.131
Special Needs			.220*	.099	.231*	.100
Siblings			.038	.027	.053*	.027
Disabled			.237**	.085	.263**	.086
Single Parent			.162*	.081	.194*	.082
SES			-.017	.039	-.128**	.041
Class Size					.008	.009
Class % Male					-.005	.004
Class % Black					-.003	.002
Class % Hispanic					-.012*	.002
Class % Reading Below Grade Level					-.011*	.002
Private School					.690	.141
School Reading %ile					-.009	.006
School Math %ile					.012*	.006
School % FR lunch					-.005*	.002

Table 1.3 – Predictors of Continuous Group Assignment in 1st grade

	Model 1		Model 2		Model 3	
Variable	Coeff.	Std Error	Coeffi.	Std. Error	Coeff.	Std Error
Intercept	5.246**	.025	.333**	.007	5.310**	.023
Prior Reading Score	1.936**	.043	.465**	.011	2.005**	.046
Prior Learning Bhvrs	.376**	.038	.053**	.010	.363**	.038
Prior Works Best	.149**	.038	.040**	.009	.143**	.038
Highest Group in K	.800**	.065	.217**	.015	.752**	.064
Lowest Group in K	-.406**	.074	.047**	.017	-.383**	.073
Ungrouped in K	.117*	.055	.093**	.013	.204**	.054
Male			-.011	.008	-.024	.034
Black			.044**	.016	-.087	.074
Hispanic			.008	.015	-.122	.065
Asian			.030	.021	.128	.087
Other Race			.037	.024	.128	.095
Non-English Home			.045**	.017	.042	.070
Special Needs			-.001	.015	-.108	.058
Siblings			-.001	.004	-.016	.015
Disabled			.007	.012	-.118*	.049
Single Parent			-.011	.011	-.120**	.045
SES			-.008	.005	.060**	.021
Class Size					-.016**	.005
Class % Male					-.001	.001
Class % Black					.005**	.001
Class % Hispanic					.008**	.001
Class % Reading Below Grade Level					.007**	.001
Private School					-.269**	.075
School Reading %ile					.001	.004
School Math %ile					-.006	.003
School % FR lunch					.003*	.001

Table 2.1 – Predictors of Highest Group Assignment in 3rd grade

	Model 1		Model 2		Model 3	
Variable	Coeff.	Std Error	Coeffi.	Std. Error	Coeff.	Std Error
Intercept	-.704**	.055	-.700**	.056	-.0681	.057
Prior Reading Score	2.740**	.149	2.652**	.157	2.992**	.165
Prior Learning Bhvrs	.481**	.090	.463**	.092	.455**	.093
Prior Works Best	.240**	.092	.246**	.093	.254**	.094
Highest Group in 1st	.930**	.101	.945**	.103	.832**	.105
Lowest Group in 1st	-.041	.143	-.030	.144	.037	.146
Ungrouped in 1st	.687**	.130	.694**	.132	.671**	.134
Male			-.019	.083	-.056	.084
Black			-.173	.146	-.549**	.182
Hispanic			.026	.141	-.165	.157
Asian			-.102	.207	-.149	.211
Other Race			-.299	.223	-.469*	.229
Non-English Home			.222	.168	.103	.176
Special Needs			-.161	.149	-.162	.152
Siblings			-.031	.037	-.055	.038
Disabled			.004	.096	.002	.098
Single Parent			.073	.108	.045	.110
SES			.149**	.049	.281**	.053
Class Size					-.028*	.012
Class % Male					.002	.004
Class % Black					.004	.003
Class % Hispanic					.002	.003
Class % Reading Below Grade Level					.012**	.003
Private School					.065	.211
School Reading %ile					-.013	.008
School Math %ile					.003	.008
School % FR lunch					.006	.003

Table 2.2 – Predictors of Lowest Group Assignment in 3rd grade

	Model 1		Model 2		Model 3	
Variable	Coeff.	Std Error	Coeffi.	Std. Error	Coeff.	Std Error
Intercept	-1.513**	.058	-1.525**	.059	-1.575**	.062
Prior Reading Score	-1.660**	.138	-1.587**	.147	-1.888**	.156
Prior Learning Bhvrs	-.566**	.093	-.534**	.095	-.534**	.096
Prior Works Best	-.044	.097	-.057	.098	-.062	.099
Highest Group in 1st	-.609**	.139	-.626**	.140	-.501**	.142
Lowest Group in 1st	.575**	.113	.557**	.114	.482**	.116
Ungrouped in 1st	-.104	.140	-.109	.141	-.129	.143
Male			-.030	.089	-.026	.090
Black			-.016	.142	.102	.177
Hispanic			-.302*	.152	-.033	.167
Asian			-.028	.243	-.033	.248
Other Race			-.038	.236	.094	.239
Non-English Home			.047	.175	.203	.183
Special Needs			.197	.135	.177	.136
Siblings			.008	.038	.023	.039
Disabled			.163	.097	.166	.098
Single Parent			-.013	.110	-.008	.111
SES			-.131*	.053	-.277**	.057
Class Size					.004	.012
Class % Male					.000	.004
Class % Black					.003	.003
Class % Hispanic					-.003	.003
Class % Reading Below Grade Level					-.008*	.003
Private School					.385	.212
School Reading %ile					.011	.008
School Math %ile					-.003	.008
School % FR lunch					-.009*	.003

Table 2.3 – Predictors of Continuous Group Assignment in 3rd grade

	Model 1		Model 2		Model 3	
Variable	Coeff.	Std Error	Coeffi.	Std. Error	Coeff.	Std Error
Intercept	5.213**	.023	5.215**	.024	5.265**	.023
Prior Reading Score	1.147**	.057	1.091**	.061	1.227**	.062
Prior Learning Bhvrs	.317**	.040	.300**	.040	.283**	.040
Prior Works Best	.068	.040	.074	.040	.075	.040
Highest Group in 1st	.457**	.051	.468**	.051	.413**	.051
Lowest Group in 1st	-.339**	.057	-.329**	.057	-.284**	.057
Ungrouped in 1st	.014	.055	.020	.055	.065	.054
Male			.005	.036	.001	.036
Black			-.016	.068	-.177*	.082
Hispanic			.105	.064	-.008	.070
Asian			.026	.094	.012	.093
Other Race			-.052	.100	-.135	.100
Non-English Home			-.012	.074	-.105	.076
Special Needs			-.097	.061	-.080	.060
Siblings			-.018	.017	-.027	.016
Disabled			-.091*	.041	-.097*	.041
Single Parent			.032	.048	.013	.048
SES			.064**	.022	.143*	.023
Class Size					-.001	.005
Class % Male					-.002*	.001
Class % Black					.000	.001
Class % Hispanic					.000	.001
Class % Reading Below Grade Level					.005**	.001
Private School					-.245**	.075
School Reading %ile					-.007*	.004
School Math %ile					.004	.003
School % FR lunch					.006**	.001

As expected, prior reading standardized test scores were by far the strongest predictor of group assignments. However, even after conditioning on prior reading scores, prior student learning behaviors were also strongly related to grouping assignments. Across models, the composite estimate of learning behaviors typically yielded a standardized effect size of about one-quarter to one-half that of prior reading achievement. Prior teacher reports that sampled students worked to the best of their abilities were also generally related to group assignments, although effect sizes were substantively smaller. This is consistent with prior research utilizing the ECLS-K dataset (Catsambis and Buttaro, 2012; Condron, 2007; Hong et al., 2012; Tach and Farkas, 2006) as well as observational studies of classrooms (Clark-Ibanez, 2005; Eder, 1981, 1983; Haller and Waterman, 1985).

Overall, individual student background characteristics had less predictive power compared to student achievement and learning behaviors. Gender disparities were negligible[5], and racial disparities were small and inconsistent across models. Students' socioeconomic status had small but statistically and substantively significant effects, particularly in 3rd grade. A one standard deviation increase in individual student SES in 3rd grade was associated with a 28.1% increase (95% confidence interval 19.5%-47.0%) in the odds of being assigned to the highest group and a 24.2% decrease (95% CI 15.2%-32.2%) in the odds of being assigned to the lowest group. There is also evidence that students that were ungrouped in prior years enjoyed slightly above-average group placements, especially in first grade.

Classroom and school contextual effects were also important. The findings suggest that while lower individual SES was associated with lower group placements, attending a school with higher proportions of students from disadvantaged families (indicated by a higher proportion of

[5] Male students are substantially more likely to be assigned to lower groups on average, but this disparity is generally a function of differences in prior achievement and learning behaviors.

students receiving federally-subsidized school lunches) was associated with *better* odds of being placed into a higher group. Similarly, higher reading achievement was strongly associated with higher reading group placements, but being part of a class with a greater proportion of students reading below grade level also increased the likelihood of a higher group placement. This highlights the importance of a student's standing relative to his or her peers. Teachers may be limited in their ability to adapt group sizes when faced with classrooms with uneven distributions of student ability. For example, if a teacher faces a classroom with a large number of students reading below grade level, she may be reluctant to create a very large bottom group, so some students are more likely to be bumped to a middle group (Eder, 1983). Students at the same level of preparation may be placed in different group levels depending on the nature of the school and classroom they attend.

The findings also indicate that the racial and gender makeup of the classrooms was influential to group assignments. While the individual effect of being male was generally insignificant in the most comprehensive models, the gender composition of the classroom is significant across all three definitions of grouping in 1st grade. The coefficients indicate that attending a classroom with a greater number of male students was associated with higher group placements. Even though gender disparities in group assignments were largely explained by other variables (namely prior achievement and learning behaviors), it remains the case that boys were disproportionately assigned to lower groups in the early grades. The presence of a larger number of boys in the classroom might mean that other students might have been "crowded out" of lower groups. On the other hand, in 3rd grade, the gender composition of the classroom was insignificant when considering highest and lowest group assignments, and associated with *lower* group placements when the continuous measure is considered. Across all models in both waves,

the effects of classroom gender composition were very small (with standardized effect sizes under .1), so it would be unwise to read too much into these findings.

Classroom racial composition appears to have been slightly more important, at least in 1st grade. Notice that across all three grouping definitions in 1st grade, African-American students were more likely to be assigned to *higher* groups when individual characteristics are controlled for (as in Model 2). This finding was surprising but not without precedence in the literature (Alexander, Cook, and McDill, 1978; Gamoran and Mare, 1989; Garet and DeLany, 1988; Jones et al., 1995). However, whenever classroom and school contextual variables are added, this advantage disappeared. Attending a classroom with a greater number of black or Hispanic classmates was associated with higher group placements, even after controlling for average classroom- and school-level reading achievement levels. This explained the group assignment advantage for black students in first grade. When classroom racial composition was allowed to modify the effect of black ethnicity at level 1, black students no longer had an advantage in classroom with an average number of black students. However, as the proportion of black students in a classroom increased, so did the probability that a black student would be assigned to a higher group. See Appendix 2 for details.

A series of stepwise regressions (in a single-level framework) explored the predictive power of the remaining variables described in Appendix 1. The effects of other student background characteristics and contextual variables were generally small and inconsistent across models. The only other variable consistently related to group assignment was students' prior math achievement test scores, suggesting that teachers used information on students' proficiencies in other subjects and evaluate them as learners as a whole when making grouping decisions. In nearly every model, prior reading achievement was the stronger predictor, followed

by the composite measure of learning behaviors. Also, in each case, prior reading achievement provided most of the predictive power. For example, consider the logistic regression models predicting highest group assignment in 1st grade. The concordance index (defined as the area under the receiver operating characteristic curve) was .795 with only reading achievement predicting group assignment, .834 with the covariates described in Model 3 in Table 1.1, and .882 in the full stepwise model with all significant covariates included.

How much mobility between group levels occurs, and what academic and behavioral characteristics predict this mobility?

There was substantial opportunity for students to move up, both during the year and between school years. Two figures below describe between-year mobility patterns. Students assigned to the lowest ability groups in kindergarten had about a 51% chance of being placed in a higher group in first grade, and similar levels of upward mobility occurred between first and third grades. However, it was also possible for students in higher groups to move down between years. It was quite common for teachers to modify group assignments *during* the school year. This occurred for about a third of students in grouped classes, and when it did, it was usually a teacher moving a student up to a higher group. Upward mobility during the school year was about five times more common than downward mobility.

A significant minority of students in all grades attended classes that did not use grouping for reading instruction. Students who were ungrouped in kindergarten but moved into grouped classrooms in 1st grade had similar placements in the highest group levels, but were slightly less likely to be assigned the lowest groups in 1st grade relative to their peers that attended grouped instruction in kindergarten (24 versus 22 percent, χ^2=5.9, p=.02). Students who were in

ungrouped classes in 1st grade had similar placements lowest-group placements in 3rd grade as their peers that attended grouped classes in 1st grade, but were more likely to be placed in the highest group in 3rd grade (42 versus 38 percent, χ^2=5.6, p=.02). Considering students who moved from grouped classes to ungrouped classes as they progressed in grade, both students in the highest and the lowest groups in kindergarten were more likely to be in ungrouped classes in 1st grade relative to their peers. Similarly, students that were in the highest group in 1st grade were more likely to be in ungrouped classes in 3rd grade (48 versus 43 percent, χ^2=14.4, p<.001). However, students in the lowest groups in 1st grade were not. It is possible that parents might have been dissatisfied with grouped instruction their children received and sought to have them placed in classrooms that did not use grouping. If so, it is interesting that movement from grouped to ungrouped classrooms is greater for students assigned to the highest groups.

Figure 2 – Ability Group Mobility Between Kindergarten and 1st Grade

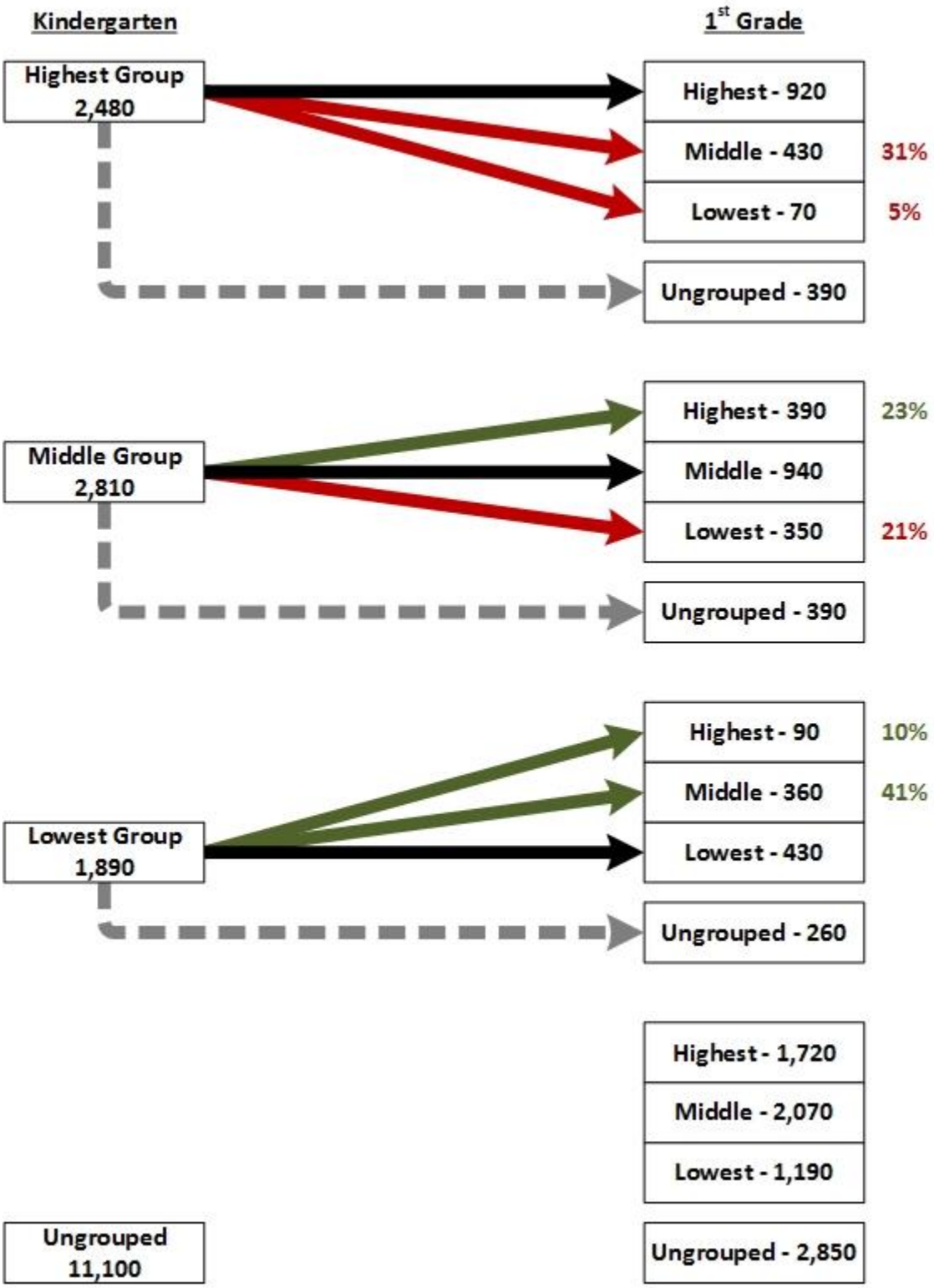
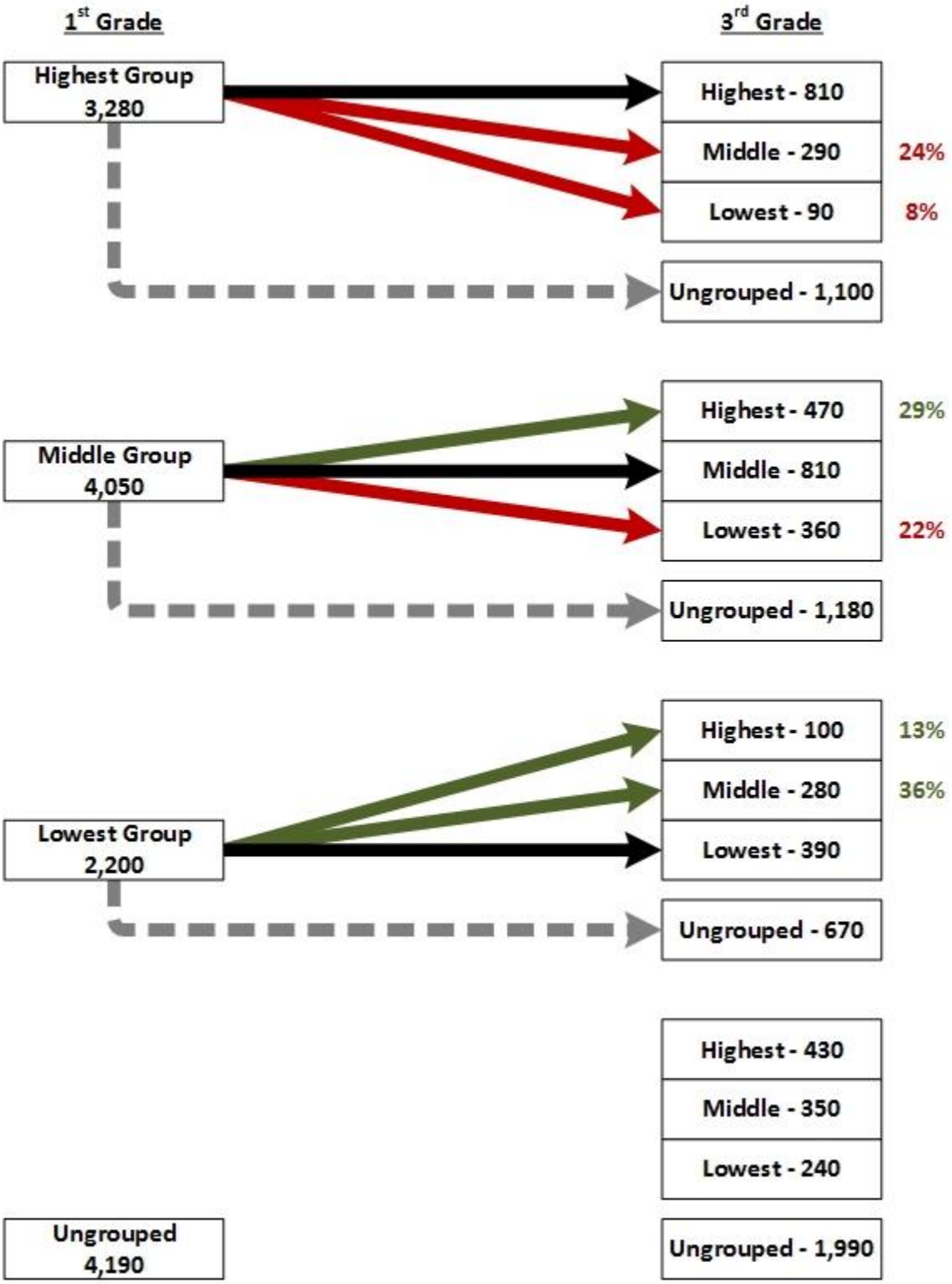


Figure 3 – Ability Group Mobility Between 1st and 3rd Grades



Regarding predictors of mobility between groups, we shift our attention to the continuous measure of group assignment. Two HLMs predicting mobility between groups (that is, the difference in group assignments between two waves of data collection) are described in Tables 3.1 and 3.2 below.[6] Findings generally mirror those describing group assignments discussed in the prior section. With the exception of prior group assignment, prior reading achievement was the most important predictor of group mobility, followed by prior learning behaviors. There was a substantial negative relationship between prior group assignments and mobility. This could be a result of the mechanical relationship between mobility and group placement; that is, the higher one's prior group placement, the more groups that one might be assigned to in the next year that would represent a downward move, and vice versa. It is also possible that the process of group assignments might be prone to error (consider the substantial overlap in the probabilities of group assignments in Figure 4 in the next section). If so, this could result in a tendency for group assignments to regress toward the mean.

[6] For this analysis, students in ungrouped classes in the prior year were assigned a prior year group measure of 5, equivalent to a middle-group placement.

Table 3.1 – Predictors of Continuous Group Mobility between Kindergarten and 1st Grades

	Model 1		Model 2		Model 3	
Variable	Coeff.	Std Error	Coeff.	Std Error	Coeff.	Std Error
Intercept	.098**	.025	.097	.024	.162**	.023
Prior Reading Score	1.910**	.044	1.919**	.046	1.981**	.046
Prior Learning Bhvrs	.354**	.038	.336**	.038	.343**	.038
Prior Works Best	.151**	.038	.146**	.038	.144**	.037
Continuous Group in K	-.797**	.011	-.797**	.011	-.808**	.011
Ungrouped in K	.016	.044	.040	.043	.110**	.042
Male			-.026	.034	-.027	.034
Black			.210**	.061	-.076	.074
Hispanic			.114	.060	-.125	.065
Asian			.148	.087	.129	.087
Other Race			.211*	.096	.126	.094
Non-English Home			.237**	.068	.050	.070
Special Needs			-.103	.058	-.101	.058
Siblings			-.007	.015	-.014	.015
Disabled			-.105	.049*	-.114*	.049
Single Parent			-.100*	.046	-.117*	.045
SES			-.014	.021	.060**	.021
Class Size					-.016*	.005
Class % Male					-.001	.001
Class % Black					.005**	.001
Class % Hispanic					.008**	.001
Class % Reading Below Grade Level					.007**	.001
Private School					-.257**	.075
School Reading %ile					.001	.004
School Math %ile					-.006	.003
School % FR lunch					.003*	.001

Table 3.2 – Predictors of Continuous Group Mobility between 1st and 3rd Grades

	Model 1		Model 2		Model 3	
Variable	Coeff.	Std Error	Coeff.	Std Error	Coeff.	Std Error
Intercept	-.103**	.022	-.101**	.022	-.062	.023
Prior Reading Score	1.105**	.055	1.047**	.058	1.175**	.059
Prior Learning Bhvrs	.292**	.038	.275**	.038	.260**	.038
Prior Works Best	.075	.038	.080*	.038	.082*	.038
Continuous Group in K	-.864**	.009	-.864**	.009	-.879**	.009
Ungrouped in K	.005	.043	.006	.043	.038	.043
Male			.010	.034	.006	.034
Black			-.017	.064	-.174*	.077
Hispanic			.124*	.061	.036	.066
Asian			.087	.089	.069	.089
Other Race			-.069	.096	-.152	.095
Non-English Home			-.013	.070	-.086	.072
Special Needs			-.134	.058	-.124*	.057
Siblings			-.029*	.016	-.037*	.015
Disabled			-.090	.039	-.095*	.039
Single Parent			.025	.046	.008	.046
SES			.061**	.021	.138**	.022
Class Size					-.001	.004
Class % Male					-.002	.001
Class % Black					-.001	.001
Class % Hispanic					-.001	.001
Class % Reading Below Grade Level					.004**	.001
Private School					-.195**	.071
School Reading %ile					-.009**	.003
School Math %ile					.005	.003
School % FR lunch					.007**	.001

Considering student background variables, higher individual student SES was associated with upward mobility across both transition periods[7]. However, it was easier to move up in a school that has greater levels of disadvantage, particularly in 3rd grade. In first grade, it was easier to move up when moving into classrooms with a greater proportion of minority students. Across both grades, it was also easier to move up when moving into classrooms with a greater number of students that are reading below grade level. Given the importance of school and classroom context, one might ask whether it was necessary for students to change schools in order to improve their chances of being assigned to higher groups. By comparing the group placement rates of students that changed schools with those of their stable peers, we can see that this was not the case. Approximately 10% of the sample changed schools between the spring Kindergarten and spring 1st grade waves, and 17% of students changed schools between spring of 1st grade and spring of 3rd grade. Students that changed schools typically ended up in lower group placements following the move. This might be due to disruptions associated with school changes or events in the lives of children that precipitated the school changes (Hanushek, Kain, and Rivken, 2004; Grigg, 2012).

[7] SES is time varying, and measured with each wave of data. It is also standardized to have a SD of approximately 1 in each wave.

Table 4 – Group assignments for mobile and stable students

Highest Group in Kindergarten			
	All	Stable Students	Changed Schools
Highest Group 1st	920 (65%)	880 (65%)	40 (56%)
Middle Group(s) 1st	430 (30%)	400 (30%)	20 (32%)
Lowest Group 1st	70 (5%)	60 (5%)	10 (12%)
Middle Group(s) Kindergarten			
	All	Stable Students	Changed Schools
Highest Group 1st	390 (23%)	370 (23%)	10 (20%)
Middle Group(s) 1st	940 (56%)	900 (56%)	40 (51%)
Lowest Group 1st	350 (21%)	330 (21%)	20 (29%)
Lowest Group in Kindergarten			
	All	Stable Students	Changed Schools
Highest Group 1st	90 (10%)	80 (10%)	10 (12%)
Middle Group(s) 1st	360 (41%)	340 (41%)	20 (41%)
Lowest Group 1st	430 (49%)	410 (49%)	20 (47%)
Ungrouped in Kindergarten			
	All	Stable Students	Changed Schools
Highest Group 1st	1720 (35%)	1620 (35%)	100 (29%)
Middle Group(s) 1st	2070 (42%)	1930 (42%)	140 (43%)
Lowest Group 1st	1190 (24%)	1100 (24%)	100 (28%)

Highest Group in 1st Grade			
	All	Stable Students	Changed Schools
Highest Group 3rd	810 (68%)	740 (69%)	80 (62%)
Middle Group(s) 3rd	290 (24%)	250 (23%)	40 (28%)
Lowest Group 3rd	90 (8%)	80 (8%)	10 (10%)
Middle Group(s) 1st Grade			
	All	Stable Students	Changed Schools
Highest Group 3rd	470 (29%)	430 (29%)	50 (23%)
Middle Group(s) 3rd	810 (49%)	700 (49%)	110 (54%)
Lowest Group 3rd	360 (22%)	320 (22%)	50 (23%)
Lowest Group in 1st Grade			
	All	Stable Students	Changed Schools
Highest Group 3rd	100 (13%)	90 (14%)	10 (11%)
Middle Group(s) 3rd	280 (36%)	240 (35%)	40 (43%)
Lowest Group 3rd	390 (50%)	340 (51%)	50 (47%)
Ungrouped in 1st Grade			
	All	Stable Students	Changed Schools
Highest Group 3rd	430 (42%)	380 (44%)	50 (33%)
Middle Group(s) 3rd	350 (34%)	270 (32%)	70 (46%)
Lowest Group 3rd	240 (24%)	210 (24%)	30 (22%)

Finally, one may question whether or not differences in the continuous measure of group assignment adequately capture the process of mobility when students transition between classrooms with different numbers of groups. Pallas et al.'s measure is adaptive to context, so a student that was in the highest group of 5 is coded with a higher value (9) than a student assigned to the higher group of 2 (7.5). This is advantageous, particularly when considering the effects of group assignment, as it reflects probably differences in the experiences of these two students. Given similar distributions of student ability within classrooms, a student in the highest group of five will likely have higher-ability peers, higher expectations from teachers relative to other students, and instruction pitched at a higher level than a student in the higher of two groups. However, when considering predictors group mobility, this is perhaps less useful because we are focused on antecedents of group assignment. A student in the highest group of five will be coded as downwardly mobile if she moves into a classroom with fewer groups, no matter how stellar her performance was in the previous year. As a specification check, I examined the subsample of students who were in the classrooms with the same number of groups in consecutive waves. There were 1360 who attended classes with the same number of groups in kindergarten and 1st grade, and 1250 who attended classes with the same number of groups in first and third grades. The tables below describe predictors of upward mobility. The outcome variable is dichotomous, coded as 1 if the student was in a higher group during the second wave than she was during the first. Upward mobility rates were about 24% from kindergarten to first grade, and 28% between first and third grades. Downward mobility rates were about 25% and 20%, respectively.

Table 5.1 – Predictors of Non-Structural Group Mobility between Kindergarten and 1st Grades

	Model 1		Model 2		Model 3	
Variable	Coeff.	Std Error	Coeff.	Std Error	Coeff.	Std Error
Intercept	-1.386**	.099	-1.427**	.103	-1.472**	.107
Prior Reading Score	1.431**	.231	1.480**	.248	1.597**	.255
Prior Learning Bhvrs	.350*	.176	.332	.180	.383*	.182
Prior Works Best	-.022	.177	-.089	.181	-.060	.183
Continuous Group in K	-.676**	.048	-.682**	.049	-.719**	.051
Male			-.162	.166	-.174	.168
Black			.019	.256	-.015	.323
Hispanic			.212	.265	-.064	.298
Asian			.850*	.362	.841*	.363
Other Race			.154	.415	.120	.414
Non-English Home			.503	.274	.299	.289
Special Needs			.161	.262	.187	.265
Siblings			-.007	.071	-.029	.072
Disabled			.057	.232	.057	.234
Single Parent			.262	.217	.286	.219
SES			.112	.104	.242*	.111
Class Size					-.051*	.024
Class % Male					-.007	.010
Class % Black					.001	.005
Class % Hispanic					.010*	.005
Class % Reading Below Grade Level					.001	.006
Private School					-1.142*	.435
School Reading %ile					.026	.015
School Math %ile					-.024	.014
School % FR lunch					.005	.005

Table 5.2 – Predictors of Non-Structural Group Mobility between 1st and 3rd Grades

	Model 1		Model 2		Model 3	
Variable	Coeff.	Std Error	Coeff.	Std Error	Coeff.	Std Error
Intercept	-1.163**	.096	-1.183**	.098	-1.22**	.102
Prior Reading Score	1.495**	.278	1.476**	.302	1.857**	.325
Prior Learning Bhvrs	.366*	.175	.260	.181	.242	.184
Prior Works Best	-.009	.177	-.003	.180	.070	.183
Continuous Group in 1st	-.643**	.047	-.660**	.049	-.713**	.052
Male			-.090	.160	-.153	.166
Black			.051	.273	-.102	.349
Hispanic			.685**	.263	.439	.295
Asian			.225	.409	-.007	.423
Other Race			.456	.428	.286	.441
Non-English Home			.201	.301	-.026	.323
Special Needs			-.176	.270	-.182	.273
Siblings			.055	.070	.057	.072
Disabled			-.302	.183	-.341	.186
Single Parent			-.088	.214	-.136	.217
SES			.275**	.094	.401**	.101
Class Size					.002	.023
Class % Male					.015	.008
Class % Black					-.003	.006
Class % Hispanic					.002	.005
Class % Reading Below Grade Level					.007	.006
Private School					-.622	.401
School Reading %ile					-.017	.014
School Math %ile					.007	.014
School % FR lunch					.009	.006

Results were generally consistent with the findings from tables 3.1 and 3.2, although fewer estimates are statistically significant. These models were estimated with considerably less power due to the additional restrictions on the sample. Prior reading achievement, prior learning behaviors, and individual-level SES all predicted upward mobility. Class- and school-level disadvantage also made upward mobility easier. Similar models with downward mobility as the outcome provided convergent results.

How does assignment to a high or low achievement group affect academic outcomes and subsequent learning behaviors?

To address the effects of group assignments, I implemented propensity score stratification to adjust for the strong selection bias associated with the treatments. Because these treatments are defined in several different ways, it was necessary to conduct each analysis separately with its own propensity model and stratification. I describe the stratification and balancing process below, using assignment to the highest group in 1st grade as an example. I then provide a summary of the results across different treatment definitions and modeling strategies.

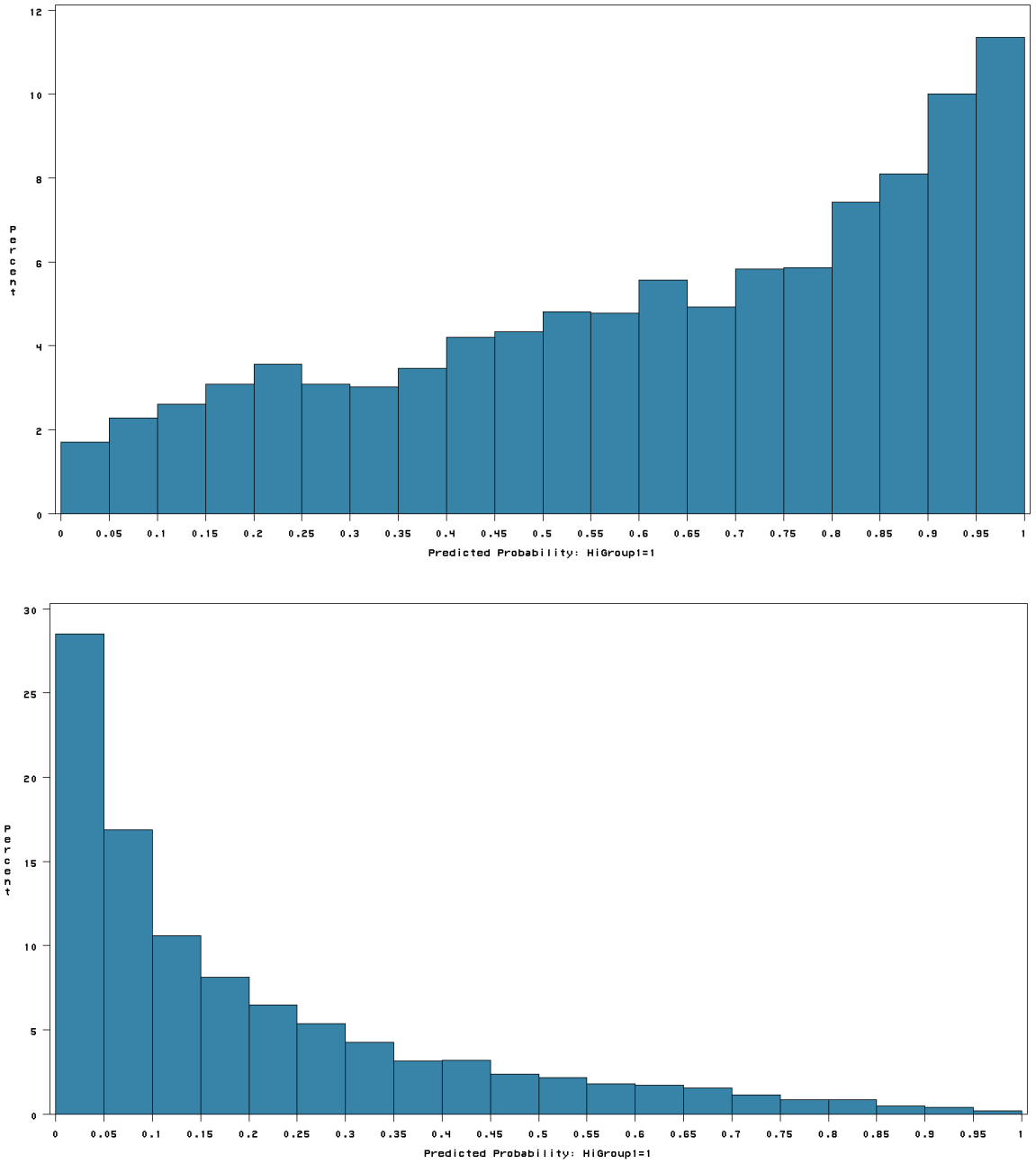
Each student's probability of being assigned to the treatment group was predicted using a large set of covariates. 81 student-level, 40 classroom-level, and 35 school-level covariates were included in the models for 1st grade (see Appendix 1).[8] These included information about prior student achievement (including both reading and math outcomes), learning behaviors, and group assignments, student background characteristics, students' home lives, classroom environments, teacher characteristics, and school contexts. Prior reading outcome measures were adjusted for measurement error to prevent bias in the results (Raudenbush and Sadoff, 2008). Because reading assessments were administered at different times during the year, the number of school

[8] Slightly fewer variables were included in the 3rd grade models due to changes in the surveys administered.

days between the beginning of the year and assessment were included in the propensity model. Because of the mechanical relationship between potential group assignments, the number of groups to which students could have been assigned was included in the model, for both the year of analysis and the prior wave.

There was generally a good deal of overlap in the distribution of propensity scores between treated and control cases. However, any students whose probabilities lay outside of the region of common support contribute no counterfactual information and were thus dropped from the analysis. This generally resulted in the loss of only a small number of students (in the case below, 40 out of 8380).

Figure 4 – Distributions of Propensity Scores for Treated (Top) and Untreated (Bottom) Students for Highest Group Assignment in First Grade



The analytic sample was then subdivided into strata until balance on all covariates was achieved on at least 95% of treatment-by-stratum interactions; imbalance was tolerated in up to 5% of these interactions because this would be expected even by chance at traditional levels of

statistical significance. However, as an additional precaution against bias, care was taken to ensure that there was no imbalance in any stratum on the propensity scores or any prior outcome variable.[9] Consider the tables below. In each, the first row shows the balance between treated (that is, students assigned to the highest group in 1st grade) and untreated students (students assigned to any other group in 1st grade). As we might expect, there were large disparities in the probability of treatment, prior reading achievement, and prior learning behaviors. Within each stratum, however, the treated and untreated students were balanced on each variable.

[9] To stratify, I wrote a program that divided the sample into five evenly-sized strata, then measured the imbalance in each variable in each stratum by running a t-test to compare treated and untreated students. The program repeated this process, dividing the dataset into six strata and repeated the balance tests. The program continued to repeat until it had divided the dataset into twenty evenly-sized strata. I selected the number of strata that provided the best balance, then manually subdivided strata that were still unbalanced on the propensity score or prior outcomes until they were balanced. Finally, I verified that the treated and untreated students were balanced in at least 95% of all treatment-by-variable interactions.

Table 6.1 – Balance of Treated and Untreated Cases on Log Propensity Score for Highest Group Assignment in 1st grade

	Treated			Untreated			
Stratum	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**t-stat**
All	2940	-.06	.074	5406	-2.30	1.37	61.49
1	11	-5.26	.79	452	-5.13	.54	-.81
2	11	-3.92	.19	453	-4.03	.21	1.81
3	6	-3.53	.09	226	-3.56	.08	1.05
4	13	-3.25	.06	219	-3.29	.07	1.97
5	17	-2.99	.13	446	-2.96	.11	-.89
6	34	-2.60	.08	430	-2.62	.09	.76
7	46	-2.30	.10	418	-2.31	.09	.11
8	24	-2.07	.04	208	-2.08	.04	1.15
9	37	-1.93	.03	195	-1.94	.04	1.34
10	78	-1.74	.07	385	-1.75	.07	.84
11	16	-1.60	.02	100	-1.60	.02	-1.67
12	22	-1.54	.02	94	-1.54	.01	.05
13	37	-1.49	.02	79	-1.49	.02	.98
14	31	-1.43	.02	85	-1.43	.02	.76
15	110	-1.28	.06	354	-1.29	.06	.25
16	151	-1.06	.06	312	-1.07	.06	.94
17	203	-.85	.06	261	-.86	.05	1.69
18	244	-.67	.05	220	-.67	.05	1.63
19	88	-.56	.02	66	-.56	.02	.19
20	93	-.50	.01	62	-.50	.01	-1.35
21	104	-.45	.01	51	-.45	.01	-.36
22	147	-.39	.02	84	-.39	.02	.28
23	172	-.32	.02	60	-.32	.02	-.30
24	377	-.22	.04	87	-.22	.04	1.44
25	420	-.12	.03	44	-.12	.03	.73
26	448	-.03	.02	15	-.04	.02	1.13

Table 6.2 – Balance of Treated and Untreated Cases on Prior Reading Score for Highest Group Assignment in 1st grade

	Treated			Untreated			
Stratum	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**t-stat**
All	2940	-.06	.074	5406	-2.30	1.37	61.49
1	11	1.67	.20	452	1.71	.24	-.62
2	11	1.87	.32	453	1.89	.25	-.31
3	6	2.05	.17	226	1.98	.23	.73
4	13	2.08	.36	219	2.02	.24	.88
5	17	2.05	.31	446	2.06	.25	-.07
6	34	2.12	.23	430	2.10	.27	.43
7	46	2.18	.23	418	2.19	.24	-.26
8	24	2.17	.35	208	2.22	.26	-.83
9	37	2.23	.27	195	2.22	.25	.37
10	78	2.22	.30	385	2.26	.25	-1.40
11	16	2.27	.24	100	2.27	.23	.10
12	22	2.33	.30	94	2.29	.24	.63
13	37	2.32	.29	79	2.32	.23	-.00
14	31	2.31	.24	85	2.29	.26	.49
15	110	2.33	.24	354	2.36	.23	-.96
16	151	2.38	.26	312	2.37	.24	.38
17	203	2.44	.25	261	2.44	.22	-.04
18	244	2.49	.25	220	2.47	.24	.85
19	88	2.58	.25	66	2.51	.24	1.81
20	93	2.59	.30	62	2.53	.26	1.30
21	104	2.58	.28	51	2.55	.21	.61
22	147	2.59	.23	84	2.64	.23	-1.76
23	172	2.63	.28	60	2.55	.24	1.88
24	377	2.67	.27	87	2.69	.28	-.75
25	420	2.79	.28	44	2.80	.30	-.32
26	448	3.11	.34	15	3.17	.40	-.67

Table 6.3 – Balance of Treated and Untreated Cases on Prior Learning Behaviors for Highest Group Assignment in 1st grade

	Treated			Untreated			
Stratum	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**t-stat**
All	2940	-.06	.074	5406	-2.30	1.37	61.49
1	11	2.34	.64	452	2.47	.60	-.71
2	11	2.49	.50	453	2.65	.62	-.87
3	6	3.14	.88	226	2.72	.61	1.62
4	13	2.62	.58	219	2.86	.63	-1.37
5	17	2.68	.64	446	2.89	.62	-1.42
6	34	2.96	.63	430	2.89	.66	.58
7	46	2.96	.60	418	3.02	.60	-.72
8	24	3.18	.68	208	3.03	.61	1.10
9	37	3.09	.64	195	3.12	.60	-.29
10	78	3.14	.57	385	3.11	.61	.43
11	16	3.40	.57	100	3.16	.52	1.64
12	22	3.32	.59	94	3.09	.59	1.66
13	37	3.08	.61	79	3.24	.63	-1.29
14	31	3.32	.60	85	3.26	.54	.55
15	110	3.34	.58	354	3.28	.58	.84
16	151	3.31	.60	312	3.29	.60	.30
17	203	3.33	.55	261	3.41	.55	-1.48
18	244	3.43	.53	220	3.46	.52	-.57
19	88	3.53	.51	66	3.43	.49	1.16
20	93	3.44	.61	62	3.41	.51	.28
21	104	3.45	.53	51	3.46	.56	-.05
22	147	3.53	.51	84	3.52	.46	.24
23	172	3.57	.45	60	3.56	.48	.13
24	377	3.55	.45	87	3.53	.50	.32
25	420	3.62	.46	44	3.64	.45	-.29
26	448	3.69	.38	15	3.59	.64	.99

Across all covariates in the propensity model, treated and untreated groups were balanced within 96.3% of strata.

Once the stratification was set, the final analytic model was a simple regression of the respective outcome variables on the treatment, a vector of dummy variables for stratum membership, and the linear propensity score. The linear propensity score (that is, the natural log of the probability of treatment) was included to reduce any residual bias that might have remained within strata, and to improve precision of estimation. Results are as follows:

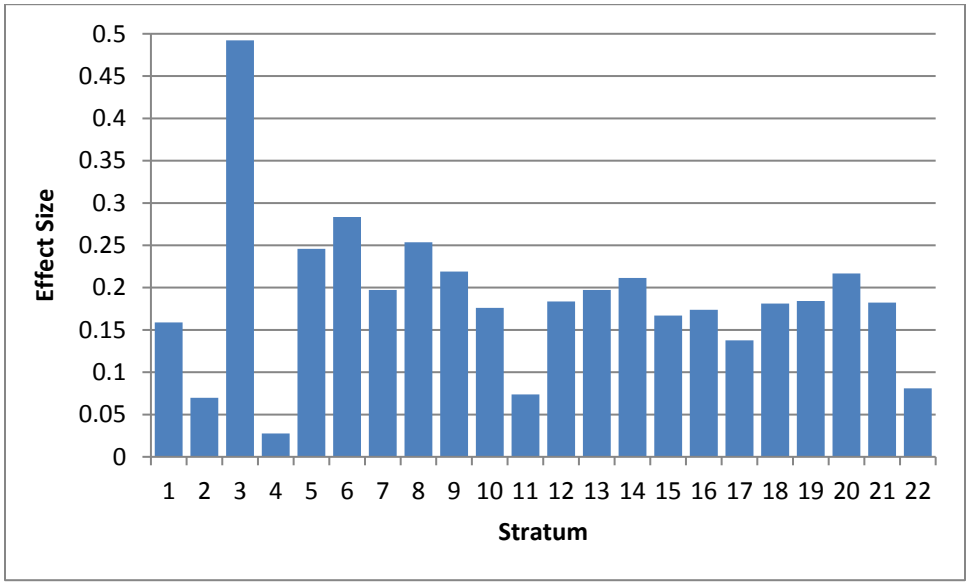
Table 7 – Propensity Score Analysis Results for Highest Group Assignment in 1st Grade

Outcome	Coefficient	Standard Error	Standardized Coefficient
Reading Achievement	.164	.009	.194
Learning Behaviors	.273	.018	.186
Works to Best Ability	.227	.018	.166

The above outcomes represent an average effect size across all strata. I also attempted to estimate effects within strata for each model by including treatment-by-strata interaction terms in the final analytic model. I did not find any evidence that the effect varied substantially across the distribution of strata. Although some strata (such as stratum 3 in Tables 6.1 through 6.3) are small and estimation is generally imprecise, there did not appear to be monotonic trends or larger effects near one or both tails.

Consider the following set of within-stratum effects:

Figure 5 – Stratum-Specific Effects for Assignment to Highest Group in 1st grade on Reading Achievement



I also conducted sensitivity analysis to explore the possibility that my findings might be biased by unobserved confounding variables. Although the propensity score models included a very comprehensive list of covariates, including power predictors such as prior reading achievement, there is always a possibility that another factor exists that is correlated with both group assignments and reading outcomes and is not adequately controlled for by covariates. I consider this by speculating that there may be an unobserved covariate that could introduce bias into my propensity score model. I regressed the outcome variables on the variables in my propensity score models, and selected one of the stronger predictors of both reading achievement and learning behaviors – students' prior math achievement scores.[10] As we noted earlier, students' math scores are also predictive of group assignments. The difference in math scores between students assigned to the highest group in 1st grade and those that were not is 5.97 points. For reading achievement, the math achievement score has a partial regression coefficient of .004.

[10] Specifically, the mathematics scale score C2MSCALE in the original dataset.

Multiplying these together yields .024, and represents the potential bias in the coefficient for group assignment in the propensity score analysis if a confounding factor similar to students' prior math scores was present but unobserved. Compared to the coefficient of .164 with a standard error of .009, this suggests that the findings for reading achievement are robust to the speculative unobserved confounding variable. For learning behaviors, the partial regression coefficient is .006, yielding a potential bias of .036. This compares to a coefficient of .273 with a standard error .018. In order to invalidate the findings observed from the propensity score model, an unobserved confounding variable would have to be many times stronger than students' prior math scores while still largely uncorrelated with the many variables in the propensity score model.

I repeated similar propensity score analyses for each dichotomous definition of grouping for each transition period. The results are described in the tables below:

Table 8.1 – Group Assignments Predicting Reading Achievement

Treatment	Control for School Variables	Coefficient	Standard Error	Standardized Coefficient
Highest Group in 1st Grade	Covariates	.164	.009	.194
	Fixed Effects	.172	.008	.207
Lowest Group in 1st Grade	Covariates	-.171	.010	-.189
	Fixed Effects	-.170	.011	-.192
Highest Group in 3rd Grade	Covariates	.093	.015	.166
	Fixed Effects	.103	.008	.190
Lowest Group in 3rd Grade	Covariates	-.102	.022	-.153
	Fixed Effects	-.100	.010	-.173

Table 8.2 – Group Assignments Predicting Learning Behaviors

Treatment	Control for School Variables	Coefficient	Standard Error	Standardized Coefficient
Highest Group in 1st Grade	Covariates	.273	.018	.186
	Fixed Effects	.262	.020	.179
Lowest Group in 1st Grade	Covariates	-.307	.020	-.187
	Fixed Effects	-.273	.023	-.169
Highest Group in 3rd Grade	Covariates	.327	.023	.236
	Fixed Effects	.307	.027	.227
Lowest Group in 3rd Grade	Covariates	-.302	.026	-.188
	Fixed Effects	-.293	.030	-.205

Table 8.3 – Group Assignments Predicting Student Working to Best of His/Her Ability

Treatment	Control for School Variables	Coefficient	Standard Error	Standardized Coefficient
Highest Group in 1st Grade	Covariates	.227	.018	.166
	Fixed Effects	.238	.020	.175
Lowest Group in 1st Grade	Covariates	-.199	.020	-.131
	Fixed Effects	-.181	.023	-.121
Highest Group in 3rd Grade	Covariates	.258	.024	.193
	Fixed Effects	.245	.027	.188
Lowest Group in 3rd Grade	Covariates	-.184	.027	-.118
	Fixed Effects	-.194	.031	-.141

All coefficients are significant with p-values less than .001. Several robustness checks were run and each yielded similar results. The above analyses were replicated with the sample restricted to subsets of students that were in ungrouped classes in Kindergarten to account for the possibility of prior treatments (kindergarten group placements) influencing outcomes in 1st grade. Replication was also done with the sample restricted to students that did not change achievement groups during the school year.

An additional concern is that teachers were responsible for both making group assignments and providing data on learning behaviors in the teacher surveys. Teachers might have been inclined to interpret the behaviors of their students in ways that validate their choices. It is also possible that they might have interpreted the behaviors of students differently based on student background characteristics such as race, gender, or social class that are correlated with group assignments (Catsambis et al., 2012, Gouldner, 1978). To explore this possibility, the propensity score analysis was repeated using the parent's report of their child's learning behaviors as the outcome. The kindergarten and first grade parent surveys included the questions for the Approaches to Learning scale. Parents' and teachers' evaluations of learning behaviors of the same children were correlated at about .26 in both years. Children behave differently at home and school, and teachers and parents may have different norms or standards for making the evaluation. However, experiences at school can affect students' behaviors at home, and if so, parents' reports should be able to tell us something important about those behaviors. In 1st grade (the only year parental reports are available for this analysis), both highest and lowest group assignments predict parental reports of learning behaviors. Standardized effect sizes were considerably smaller than when teacher reports are analyzed – .044 and -.064 respectively – but remained statistically significant with $p<.001$.

While I assumed that the highest and lowest groups were key salient positions across classrooms, limiting the analysis to them ignores the possibility of being assigned to other higher or lower groups. These possibilities will vary in classrooms with different numbers of groups. As a specification check, I also conducted propensity score analysis with treatment defined as being placed above different cut points in the distribution of continuously-defined group assignments. I used cut points of 4.9 (which would define treatment as assignment to the middle

group groups and up across all classrooms), 5 (which would include students assigned above any middle group), and 6.6 (which would include only students in the highest two groups in classrooms with five or more, and the highest group in all other classrooms).

Table 9 – Propensity Score Analysis Results with Treatment Defined by Continuous Group Placements

Treatment	Outcome	Coefficient	Standard Error	Standardized Coefficient
Continuous Group > 4.9	Reading Achievement	.165	.008	.201
Continuous Group > 4.9	Learning Behaviors	.320	.017	.218
Continuous Group > 4.9	Works Best	.196	.017	.144
Continuous Group > 5	Reading Achievement	.165	.008	.201
Continuous Group > 5	Learning Behaviors	.320	.017	.217
Continuous Group > 5	Works Best	.197	.017	.145
Continuous Group > 6.6	Reading Achievement	.173	.008	.219
Continuous Group > 6.6	Learning Behaviors	.300	.016	.213
Continuous Group > 6.6	Works Best	.231	.016	.178

These findings were convergent with those that focused on group assignments to the highest and lowest groups. They also suggest that assignment to the highest groups in the class had an even stronger effect than assignment to groups at or somewhat above the mean, providing additional support for the Matthew effects hypothesis.

The development of learning behaviors and academic outcomes induced by group placements may persist only as long as students remain in grouped settings, or they may persist even after a year of grouped instruction is concluded. To test this, I examined the relationship between group assignments in one wave on outcomes in the next. The effects of group assignments in 1st grade on 3rd grade outcomes are described below. The sample includes all students that were in grouped classes in 1st grade that also had outcome information in 3rd grade, regardless grouping status or assignment in 3rd grade. Unfortunately, changes to the teacher

survey over time mean that grouping information is not available after 3rd grade. However, the data do include reading test outcomes as well as the Approaches to Learning scale in the 5th grade wave, so I consider the persistent association between 3rd grade group assignment and 5th grade outcomes.

Table 10.1 – Group Assignments Predicting Reading Achievement in Subsequent Waves

Treatment	Outcome Year	Coefficient	Standard Error	Standardized Coefficient
Highest Group in 1st Grade	3rd Grade	.077	.008	.131
Lowest Group in 1st Grade	3rd Grade	-.102	.009	-.155
Highest Group in 3rd Grade	5th Grade	.035	.010	.062
Lowest Group in 3rd Grade	5th Grade	-.136	.013	-.201

Table 10.2 – Group Assignments Predicting Learning Behaviors in Subsequent Waves

Treatment	Outcome Year	Coefficient	Standard Error	Standardized Coefficient
Highest Group in 1st Grade	3rd Grade	.142	.022	.102
Lowest Group in 1st Grade	3rd Grade	-.115	.024	-.073
Highest Group in 3rd Grade	5th Grade	.066	.030	.047
Lowest Group in 3rd Grade	5th Grade	-.196	.033	-.119

There is substantial evidence that assignment to a higher group generally predicted improved academic and behavioral outcomes even two years after the end of instruction in a particular grouped setting. Efforts were made to calculate the interaction effects between group assignments in 1st and 3rd grades, but the data could not be balanced on all covariates across four treatment subpopulations (representing four possible patterns of treatment for each binary treatment over two years). However, analyses that minimized data imbalance (achieving balance on 85-90% of covariate-by-treatment interactions) suggested that interaction effects were not particularly large.

Discussion and Conclusion

The motivating question behind this project is "Can you work your way up?" – that is – can students that are assigned to low achievement groups early in their educational careers work hard and advance to into higher groups as they move through school. The findings indicate that the answer to this question is "Yes," at least in the early years of elementary education when within-class grouping is common. There was substantial opportunity for students to move up, both during the year and between school years. Upward mobility was substantially more common than downward. Moreover, both higher group assignments in general and upward mobility were predicted by students demonstrating positive learning behaviors, even after controlling for achievement. In fact, aside from prior reading achievement, teacher reports of student learning behaviors were the strongest predictors of group assignments in nearly every model considered. Students that demonstrated good work habits were more likely to move into higher groups as they progressed through school. This has been documented in small-scale classroom observation studies, but this chapter provides the evidence that this practice has been widespread in many classrooms across the country.

Unfortunately, the findings also indicate that despite opportunities for advancement, there are still substantial reasons for concern. I followed students' academic progress as they moved through school and experienced different group placements. In both first and third grades, students that had moved up from a lower group in a prior year were nonetheless behind their stable high-ranking peers at the end of the year. The gap averaged around one-half a standard deviation in reading scores – a very large difference. However, it appears that students can work to close these gaps over time. Students that moved up into a higher group between kindergarten and first grade were able to somewhat reduce (but not close) this gap by the end of third grade.

Even more concerning is evidence that suggested a student's socioeconomic background, relative to that of his or her peers, could influence the group assignment that a student received, even after controlling for prior achievement and learning behaviors. In contrast, racial disparities in group assignments were small and generally not statistically significant. This is consistent with most prior research that has found that, although minority students are disproportionately assigned to lower groups, the disparities in assignment are typically explained by differences in prior achievement rather than overt racial bias. Indeed, in some cases minority students have experienced higher group or track assignments after accounting for differences in prior achievement and other covariates (Alexander, Cook, and McDill, 1978; Gamoran and Mare, 1989; Garet and DeLany, 1988; Jones et al., 1995). The current findings are also convergent with, and perhaps might partially explain, evidence that gaps in achievement between children from wealthy families and those of poor families has been increasing in recent decades even as black-white achievement gaps have declined (Reardon, 2011).

Finally, an important argument in the literature on this topic is that group assignment can reinforce existing differences between students. I have found evidence consistent with this Matthew effects hypothesis. Students assigned to the highest group learned more than they would have if they had been assigned to a lower group, and students assigned to the lowest group learned less than they would have if they had been assigned to a higher one. Moreover, group assignment appears to have reinforced existing differences in learning behaviors. Students assigned to the highest groups also showed improvements in their learning over time, whereas students assigned to lower groups showed worsening behaviors. The effect sizes are substantial, and the treatments had the potential to be repeated over multiple years. Students could indeed work their way up, but something about students' experiences in low-ranking groups was

discouraging good learning behaviors. There are several potential explanations for this, including differences in instruction, the influences of different peer groups, and low teacher expectations for students in low-ranking groups. Unfortunately, I cannot discern between these mechanisms with the current data.

As is the case in nearly all modern research on achievement grouping, these analyses are based on observational data rather than experimental manipulation of treatments. While this has the advantage of examining the outcomes of educational practices as they are implemented in the real world, it limits my ability to make causal claims regarding the effects identified in the analysis. While the sensitivity analysis suggests that it is unlikely, I cannot eliminate the possibility that unmeasured covariates may confound my analysis. Furthermore, the stable unit treatment value assumption is likely violated. Grouping can be implemented in many different ways, with different numbers, sizes, and compositions of groups, and can be implemented in different combinations with other instructional practices. Lacking information on most of these factors, I reduce this complexity into a simple set of two dichotomous and one continuous treatments that do not capture all of these differences. It is also plausible that the outcomes of students might be affected by the treatment assignments of other students. For example, a student on the margin between two groups might be assigned to a lower group if the higher one already has the maximum number of students the teacher prefers in a single group (Eder, 1983).

Nonetheless, these findings provide substantial evidence that group assignment has important consequences for both student academic achievement and learning outcomes. Students assigned to higher groups experienced substantial positive changes in development as learners, and students assigned to lower groups experienced depressed development, even conditioning on prior achievement and learning behaviors. In many cases these changes

persisted after two years. This is consistent with prior literature that has found differential effects of grouping depending on whether students were assigned to higher or lower groups. It is important to reiterate here, however, that the current study has not considered what would have happened to students if they were never grouped. The implication is therefore not that grouping necessarily increases inequality in outcomes, but rather that group assignments have consequences for the development of students as learners.

Eliminating or reducing the practice of achievement grouping may not be possible, or perhaps even desired. Teachers find it easier to instruct groups of students with similar levels of preparation, and it allows some level of customization of instruction (Ireson and Hallam, 2001). It may be possible to improve instructional practices using grouping in a way that targets instruction effectively and mitigates negative effects on lower-achieving students (Hallinan, 1994). The recent randomized experiment described in Duflo et al. (2008) provided a proof-of-concept that grouping can provide learning benefits to all students, including those assigned to lower groups. But a stark contrast exists between instruction in an experimental context and the day-to-day realities of classrooms. Finley (1984) documented teachers reporting that they prefer instructional activities that emphasize creativity and higher-order thinking skills, but elect to focus on basic instruction after students in lower groups or classes fail to respond them. Evidence that typical learning experiences in low groups actually discourages learning-oriented behaviors reinforces the need to provide stimulating and challenging instruction to all students, *especially* to those that have displayed weaker prior achievement and learning behaviors.

Chapter 3 – Between-class grouping in secondary schools

Introduction

While within-class grouping for instruction in certain subjects is the modal practice in the early elementary grades, students are typically sorted between classes at the secondary level in the United States. This practice is generally referred to as "tracking," although the term is not without dispute. Some authors reserve the term for a rigid, comprehensive organization of schools where students are assigned to courses in most or all subjects on the basis of a generalized assessment of a student's aptitude or expected postsecondary destination. This arrangement was common in the middle of the 20th century, but has generally been replaced by more flexible systems that allow students to take courses of different levels across core subjects (Gamoran and Weinstein, 1988; Loveless, 1999; Lucas and Berends, 2002). Lucas (1999) described what he terms the "unremarked revolution" during the 1960's and 1970's that saw the widespread dismantling of over-arching tracking policies within schools.

In recent decades, students in a typical comprehensive high school would have some amount of choice among several vertically-ordered levels of classes in core subjects as well as different types of elective courses. These choices are often restricted by requirements such as minimum grades or test scores, recommendation from prior teachers, or the completion of pre-requisite classes. While rarely called "tracks," students are often presented with options for sets of courses designed for different post-secondary destinations. Even schools that emphasize "college-prep for all" students, there are generally choices of classes for students with different levels of preparation or ability. In the data analyzed in this chapter, 78% of 12th grade students reported that they chose their "high school program," either alone or in consultation with others. These two elements – student choice and formal assignment criteria – create important

differences in the process by which students are assigned to between-class groups versus the within-class grouping practices discussed in the prior chapter. Whereas within-class reading group assignments are made wholly by teachers (although they can sometimes be influenced by assertive parents), student enrollments in different courses in high schools are influenced by a number of actors: students, teachers, counselors, and schools as institutions. How these influences might change the outcomes of assignments or the amount of mobility between tracks is theoretically ambiguous. For example, greater student choice can create important opportunities for motivated students to access more demanding coursework regardless of their prior assignments. However, students will often make choices that reflect and recreate existing systems of stratification and segregation (Burris, 2014; Gamoran and Weinstein, 1998; Heck, Price, and Thomas, 2004). Students may choose courses to be with homophilous friends. Ethnic minority students have expressed reluctance to join honors classes that enroll mostly white students who might not perceive them as an equal (Yonezawa, Wells, and Serna, 2002). Karolyn Tyson, in her study of "racialized" tracking in high schools, explained that "without prompting, many mentioned being the 'only black' student in their advanced classes... Keisha, who attended Garden Grove High School (54% white; 32% black), complained of being the 'only black girl in a sea' of 'white people' in her advanced classes" (Tyson, 2011).

Teachers and counselors can recognize talented students, and will often encourage them to move up into higher tracks, or resist efforts to move down. However, they can also resist the efforts of students (and their parents) to enroll in more demanding courses. There is a particular concern that school actors discourage the upward mobility of disadvantaged and minority students. Lewis and Diamond (2015) document reports that, while predominantly middle-class white students were "automatically" placed in high track classes, black and Latino students were

regularly assigned to lower tracks. When the parents of minority students attempted to advocate for higher placements for their children, they encountered resistance from school personnel. Requests for courses changes were generally eventually honored, but were more likely to require persistence on the part of parents of color. Middle-class parents (who are more likely to be college-educated themselves) are also more likely to understand the consequences of course placements in the first place, that they can be influenced by advocacy, who to contact at their children's schools, and how to advocate successfully for desired changes (Delany, 1991). Yonezawa, Wells, and Serna (2002) found that, even in schools instituting detracking reforms, teachers and counselors acted as gatekeepers to dissuade or prevent some students from enrolling in advanced classes. Additionally, numerous researchers have identified the practice of teachers pushing uncooperative or disruptive students into lower track classes, even if they possessed the academic skills to succeed in more advanced classes (Oakes, 2005; Finley, 1984; Lewis and Diamond, 2015).

The organization of grouping between classes also imposes certain logistical limits on both group assignments and the potential for mobility. While teachers have the option of forming within-class groups of any size they desire, schools face limits on the size of course enrollments and the increments in which they can be changed. Limits on physical space in classrooms and norms of fairness for teachers and students constrain limits on class sizes. Schools generally will not have the resources to make substantial adjustments to teaching staff each year to accommodate student needs. Marginal students may be "crowded out" of high-track classes if there are not enough spaces in those classes. The opposite can occur as well, although schools might be more reluctant to place a student in a class perceived as too difficult rather than easy for the student. Delany (1991), in a close study of the process of course assignment and

schedule building, summarized: "In effect, the matches of students and courses became intelligent adaptations by the participants to the constraints placed on them. The matches are often the solutions to problems unrelated to students' educational needs."

Logistical constraints can not only affect idiosyncratic assignment of individual students, but shape the larger structure of course programs. A particular concern is math instruction. Due to the sequential nature of instruction, with more advanced concepts requiring mastery of prior techniques, math is generally the subject that is most heavily tracked. Furthermore, advanced mathematics courses require specialized training to teach effectively, and schools generally have a limited number of instructors that can teach those classes. Unless the school is unusually well-resourced, advanced math courses are generally only taught during certain class periods. Because many (though not all) students that enroll in the highest-level math courses will also wish to enroll in the highest level of courses available in other subjects, schools will often build their schedule to accommodate this common pattern. Thus, while it may be possible in theory for a student to take courses of different levels in different subjects, scheduling logistics can make it difficult to do so. The result is that, while schools rarely enforce formal tracking policies as in the past, logistical constraints and student choice create *de facto* tracking environments in many schools. In the data presented in this chapter, 81% of 10th grade students identified the "high school program" (academic/college-prep, general, or vocational) to which they belonged. Similarly, only about 13 percent of principal reports indicated at all students were placed in a single program (7.5% reported that all students were in an academic/college-prep program, and 5.5% indicated all students were enrolled in a general program).

Research Questions

My inquiry in this chapter progresses much the same way as in the previous chapter. The goal is to describe how students are sorted into tracks as the secondary level, how track assignments affect students' learning behaviors, and whether students with improving achievement and learning behaviors can "work their way up" into the highest track before graduation. I am also interested in whether and how the different training students experience between tracks affects their post-secondary education and labor market outcomes after high school. Specifically:

1. What student background, academic, and learning behaviors predict students' track assignments?
2. How much mobility between tracks occurs, and what academic and behavioral characteristics predict this mobility?
3. How does assignment to different tracks affect academic outcomes and subsequent learning behaviors?
4. How do track assignments in high school affect success in the labor market?

Data and Methods

Data for this chapter come from the National Education Longitudinal Study of 1988 (NELS). A nationally-representative sample of eighth-graders was first surveyed in the spring of 1988. A sample of these respondents was then resurveyed through four follow-ups in 1990, 1992, 1994, and 2000. On the questionnaires, students reported on a range of topics including: school, work, and home experiences; educational resources and support; the role in education of their parents and peers; neighborhood characteristics; educational and occupational aspirations;

and other student perceptions. For the first three waves of data collection (when most were eighth-graders, sophomores, or seniors), achievement tests in reading, mathematics, and other subjects were administered in addition to the questionnaires. Teachers were surveyed and asked questions about themselves and their perceptions of sample students. Administrators provided information about schools the sampled students attended. The initial sample included 12,140 students that began the study in 1476 high schools in 10th grade.

The primary independent variables of interests – students' ability groups and tracks – come from student survey reports. In 10th and 12th grades, some teachers also provided information as to the level of class they taught that sampled students were enrolled in, but the question was not asked of all teachers, and the sampled teachers taught different subjects. Because I am interested in change over time, the student self-reports were the only data available consistently across the first three waves. In 8th grade, students reported their "ability group" in their math class, with the options of high, middle, low, "we aren't grouped," and "I don't know." In 10th and 12th grade, students were asked which best described their current "program." The options included: "general high school program," "college prep, academic, or specialized academic (such as Science or Math)," or various options under the heading "vocational, technical, or business and career." These options included fields such as agricultural, health, or office occupations. Students could also choose "other" or "I don't know." 81% of 10th grade students identified as being part of one of the first three programs. The remaining 19% who indicated "other" or that they did not know, or did not respond to the question, were omitted from the analysis.

Outcome variables of interest include students' math and reading test scores from 10th and 12th grades, reports of student learning behaviors, and employment outcomes following high

school. As in the prior chapter, I used the IRT Theta scores representing the underlying math and reading abilities of students based on their responses on the standardized tests. As for learning behaviors, one advantage of the NELS data over the ECLS-K is that it includes reports of from both students and teachers. For the 8th grade, students were asked:

- How often do you cut or skip classes?
- How many times were you late for school over the past four weeks?
- How often do you come to class and find yourself without these things?
  - (a) Pencil or paper (when needed)
  - (b) Books (when needed)
  - (c) Your homework done (when assigned)

Students provided ordinal responses indicating frequency, and these were averaged to form the control index of learning behaviors for the 8th grade. Also in 8th grade, teachers were asked whether sampled students:

- Consistently performs below ability?
- Rarely completes homework?
- Is frequently absent?
- Is frequently tardy?
- Is consistently inattentive in class?

Teachers responded "yes," "no," or "don't know" to each item. Responses of "don't know" were set to missing. The index of teacher-reported learning behaviors in 8th grade was the average of these items.

In 10th grade, students were asked to how often they felt it was "OK to..."

- Be late for school?
- Cut a couple of classes?
- Skip school for a whole day?
- Cheat on tests?
- Copy someone else's homework?

And questions regarding preparation for class:

- How often do you come to class without these things?
  (a) Pencil or paper
  (b) Books
  (c) Your homework done

Students responded with ordinal responses on a 1-4 scale indicating frequency. These were reverse coded and averaged to form the 10th grade index of student-reported learning behaviors. Also in 10th grade, two different teachers that taught sampled students were surveyed, and both provided reports of learning behaviors. Teachers were asked:

- How often does this student complete homework assignments?
- How often is this student absent?
- How often is this student tardy?
- How often is this student attentive in class?

Teachers responded on a 1-5 scale indicating frequency, or "don't know" which I coded as missing. Responses were reverse-coded as necessary, and all eight responses (four from each teacher) were averaged to form the teacher-reported index of learning behaviors for 10th grade.

Finally, in 12th grade students were asked specifically about their current or most recent math class. Students were asked how often did they:

- Pay attention in class?
- Complete work on time?
- Do more work than was required?
- Participate actively in class?

Students responded on a 1-5 scale indicating frequency. These were averaged to form the student-reported index of learning behaviors for 12th grade. The survey items teacher-reported index of learning behaviors were identical to those described earlier for 10th grade, except only a single teacher was surveyed for each student.

Details regarding the indices of learning behaviors, including item correlations, are included in Appendix 4. Cronbach's alpha reliabilities range from .63 to .82.

To address selection bias, I used a set of 30 student-level and 19 school-level covariates from the student, parent, and school administrator surveys to form the basis of my propensity models. I chose variables that provided information about students that could potentially affect either the probability of receiving the treatments (i.e. the probability of being assigned to one track versus another) or the outcomes I consider. I did not include any covariates that could have been caused by the treatments. Descriptive statistics are detailed in Appendix 3. They include student demographic background, information about student's home lives, information about their relationship with their parents, parents' involvement in education, and data on school resources and composition. As in the prior chapter, imputation was used to address missing data, except for information on treatments (student tracks). A ridge prior with two degrees of freedom was used to alleviate multicollinearity among the covariates and allow the imputation models to converge. Findings below represent estimates from a single imputed dataset analyzed in HLM. Analysis proceeded along the same general lines as the prior chapter considering within-class grouping in primary schools. First, I considered what student background, academic, and behavioral characteristics predicted students track locations. Second, I examined mobility between tracks over time. Finally, I used propensity score stratification to examine the effects of track assignments had on academic outcomes, learning behaviors, and labor market outcomes following high school.

Track Assignments

Student choice versus school prerogatives

While the ECLS-K held no information about the process by which teachers made group assignments, the NELS surveys asked students about their involvement in the track selection

process. In the 12th grade student survey, immediately after asking respondents to identify their current high school program, an item asked "How did you get into this program?" [1] Table 11 below describes how students in the various tracks described the process. Note that students could select more than one option.

Table 11 – Student Reports of Track Selection Process

Student Response	Overall	Academic	General	Vocational
(a) I was assigned.	32.7%	29.6%	42.1%	14.8%
(b) I chose it after talking with my counselor or teacher	42.0%	48.8%	34.3%	45.5%
(c) I chose it after talking with my parents	38.7%	49.1%	28.2%	38.4%
(d) I chose it after talking with my friends	23.1%	27.5%	16.7%	27.9%
(e) I chose it myself – did not consult anyone	26.1%	24.3%	23.5%	39.0%
(f) This is the only program in my school	13.7%	11.2%	21.7%	2.6%

Students in the general track were more likely to report that they were assigned, while students in the vocational tracks were more likely to report having chosen their program. More students in the academic track indicated that they consulted with adults such as parents, teachers, and counselors before making their decision to pursue that program. I ran a logistic regression analysis to explore what types of students were afforded the opportunity to choose their program.[2] The results suggested that student input into the selection process was generally associated with prior student advantage. Students with higher 8th-grade test scores or who were in high ability groups in 8th grade were more likely to report that they had chosen their program. Asian students were more likely to report having chosen, while black and Hispanic students were less likely. Socioeconomic status and gender were not significantly associated with choosing. It is important to note that student choice and school procedural assignments are not mutually

[1] Students who dropped out prior to the administration of this survey were offered a different instrument and were not asked this question.
[2] Specifically, I created an outcome variable coded 1 if a student answered "Yes" to any of the five responses beginning "I chose..." in Table 11. I regressed this outcome against student background characteristics.

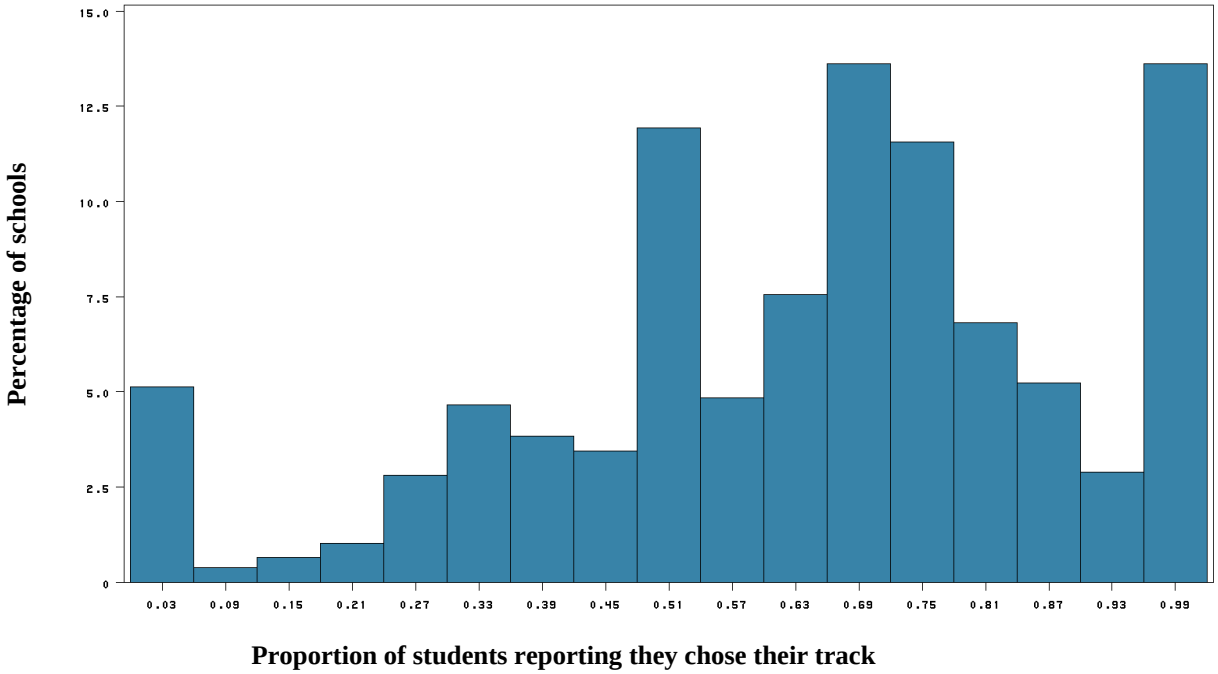
exclusive. Approximately 22% of students (including about 23% of students in academic and general tracks but only two percent of students in vocational tracks) reported both that they had been assigned to the program *and* that they had chosen it. This is consistent with processes described by Delany (1991) and Yonezawa, Wells, and Serna (2002) where student choices were both accommodated and adapted to the institutional requirements of schools. Table 12 below describes three categories of students: those that only indicated that they had chosen their program (selecting "Yes" to options B, C, D, or E to the question in table one, and "No" to all other responses), those that only indicated that had been assigned (selected "Yes to options A or F, and "No" to all other responses), and those who indicate both that they had been assigned and had chosen (selected "Yes" on at least one of B, C, D, or E *and* "Yes" on either A or F). These categories are mutually exclusive.

Table 12 – Student Reports of Choice in Track Selection Process

Assignment Procedure	Overall	Academic	General	Vocational
Student chose track	57.8%	61.6%	42.6%	83.5%
Student was assigned to track	21.3%	15.0%	34.0%	5.7%
Student chose and was also assigned	21.6%	23.3%	23.4%	10.7%

It is also clear that assignment procedures could vary from student to student within schools. Over a thousand schools in the study had at least two students that answered questions about their track assignment on the survey. Among them, a majority (62%) had at least one student that only indicated that he or she was assigned *and* at least one other student that indicated that he or she was not assigned but chose their selected program. Figure 6 below is a histogram showing the distribution of school-level means of students reporting only that they chose their programs and were not assigned. The sample size within schools for this analysis ranges from 2 to 37 with a mean of about 9.

Figure 6 – School Means of Student Reports of Track Choice



Overall, the data depict a clear pattern where, when given a choice, students were more likely to have selected either the academic track or the vocational track by 12th grade. Students in the general track were more likely to have been assigned (relative to students in the other programs), although it is important to note that a majority of these students still reported that they were involved in the selection process. Also, the 12th grade school administrator survey included a question about assignment criteria for the vocational track. The reliance on student self-selection was nearly unanimous, although these choices were typically made in conjunction with staff approval or other criteria set by the school.

Table 13 – Principal Reports of Assignment Criteria for Vocational Tracks

Student selection[α]	96.6%
Teacher referral	79.1%
Grade point average	21.2%
Counselor referral	85.6%
Other	47.5%

[α] Indicates that students self-selected into vocational programs.

Predictors of Track Assignments

As discussed in the introduction, both student choice and institutional influences can guide different types of students into different track locations. While most prior research on tracking has examined student background characteristics, very few studies have examined how students learning behaviors change their track destinations (Carbonaro, 2005). I ran a series of hierarchical multinomial logistic regressions aimed at describing what student and school characteristics predict students' enrollments in academic and vocational program. Results are detailed in the tables below. In each table, Model 1 includes prior group assignments, prior achievement, and prior learning behaviors. Model 2 adds student-level covariates. Model 3 adds school-level covariates modifying the intercepts only. Each model includes a random intercept and fixed slopes for all covariates[3], and all covariates are centered around their grand means. The reference group is the general track. The full model for table 14.1 is included in Appendix 5 as an example. Students are nested in the schools they attended at the time of the outcome.

[3] Models with random slopes (not reported) were difficult to estimate and indicated that there was no appreciable variance between schools in slopes for key covariates.

Table 14.1 – Predictors of Academic Track Enrollment in 10th Grade

	Model 1		Model 2		Model 3	
Variable	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Intercept	-.359**	.043	-.388**	.043	-.370**	.043
Prior Math	.043**	.005	.035**	.005	.036**	.005
Prior Reading	.038**	.004	.033**	.004	.033**	.004
Prior LB-Student report	.268**	.061	.260**	.063	.266**	.063
Prior LB-Teacher report	.949**	.133	.857**	.135	.870**	.136
High Group in 8th	.536**	.066	.411**	.069	.404**	.069
Low Group in 8th	-.248	.133	-.227	.134	-.225	.134
Ungrouped in 8th	-.062	.080	-.041	.081	-.034	.081
Missing Group Info	.059	.117	-.019	.120	-.036	.121
Male			.051	.054	.051	.054
Black			.609**	.109	.500**	.115
Hispanic			.166	.112	.056	.116
Asian			.179	.123	.134	.124
Siblings			-.064**	.018	-.064**	.018
Non-English Home			-.070	.046	-.048	.047
Gifted			.461**	.070	.444**	.070
Prior Retained			-.338**	.093	-.337**	.093
Discuss			.091*	.042	.090*	.042
Moved			-.095	.075	-.097	.075
Married			-.108	.123	-.108	.123
Divorced			-.049	.109	-.053	.109
SES			.236**	.041	.230**	.042
%Single Parents					.070	.073
% LEP					.027	.039
% FR Lunch					-.001	.003
% White					-.005*	.002
% Parent Volunteer					.004	.003
Magnet					.097	.188
Catholic					-.096	.230
Public Choice					-.411**	.125

Table 14.2 – Predictors of Academic Track Enrollment in 12th Grade

	Model 1		Model 2		Model 3	
Variable	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Intercept	.032	.036	.014	.036	.030	.036
Prior Math	.058**	.004	.052**	.005	.052**	.005
Prior Reading	.027**	.004	.022**	.004	.022**	.004
Prior LB-Student report	.202**	.059	.195**	.061	.203**	.061
Prior LB-Teacher report	.470**	.055	.447**	.056	.447**	.056
Academic Track in 10th	2.165**	.065	2.118**	.065	2.112**	.065
Vocational Track in 10th	.475**	.114	.508**	.115	.522**	.115
Missing Track in 10th	.534**	.075	.510**	.077	.503**	.077
Male			-.084	.055	-.086	.055
Black			.502**	.103	.449**	.110
Hispanic			.122	.106	.108	.111
Asian			.172	.121	.194	.123
Siblings			-.043*	.018	-.043*	.018
Non-English Home			-.032	.045	-.022	.045
Gifted			.220**	.072	.228**	.072
Discuss			.102*	.042	.100*	.042
Moved			.085	.075	.098	.075
Married			-.141	.119	-.134	.119
Divorced			-.011	.110	-.019	.110
SES			.325**	.041	.318**	.042
%Single Parents					.001	.047
% LEP					-.065	.034
% FR Lunch					.002	.002
% White					-.001	.002
% Parent Volunteer					-.005	.003
Magnet					.060	.129
Catholic					.664**	.153
Public Choice					-.230**	.082

Table 15.1 – Predictors of Vocational Track Enrollment in 10th Grade

	Model 1		Model 2		Model 3	
Variable	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Intercept	-1.714**	.048	-1.804**	.051	-1.826**	.052
Prior Math	-.046**	.007	-.036**	.007	-.035**	.007
Prior Reading	-.036**	.006	-.023**	.006	-.023**	.006
Prior LB-Student report	.079	.073	.084	.076	.094	.076
Prior LB-Teacher report	-.482**	.143	-.320*	.147	-.305*	.147
High Group in 8th	.074	.105	.016	.108	.012	.108
Low Group in 8th	-.197	.139	-.178	.142	-.170	.142
Ungrouped in 8th	-.041	.099	-.084	.101	-.067	.101
Missing Group Info	.188	.126	.138	.133	.143	.133
Male			.407**	.076	.407**	.076
Black			.528**	.119	.484**	.130
Hispanic			-.065	.139	-.050	.146
Asian			.004	.193	.016	.195
Siblings			.014	.023	.015	.023
Non-English Home			-.049	.062	-.061	.063
Gifted			.095	.111	.087	.112
Prior Retained			.180	.096	.191*	.097
Discuss			.033	.054	.037	.054
Moved			-.097	.101	-.096	.101
Married			-.118	.160	-.125	.160
Divorced			.189	.135	.188	.135
SES			-.543**	.058	-.511**	.059
%Single Parents					.187*	.075
% LEP					-.041	.039
% FR Lunch					.002	.003
% White					.001	.002
% Parent Volunteer					-.008*	.004
Magnet					.303	.183
Catholic					-.594	.334
Public Choice					-.286*	.125

Table 15.2 – Predictors of Vocational Track Enrollment in 12th Grade

	Model 1		Model 2		Model 3	
Variable	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Intercept	-1.240**	.048	-1.296**	.049	-1.340	.052
Prior Math	-.011*	.006	-.004	.006	-.003	.006
Prior Reading	-.011*	.005	-.011*	.005	-.011*	.005
Prior LB-Student report	.184**	.070	.138[β]	.072	.132[γ]	.072
Prior LB-Teacher report	-.114	.063	-.099	.064	-.093	.064
Academic Track in 10th	.357**	.105	.406**	.106	.394**	.106
Vocational Track in 10th	1.586**	.093	1.569**	.094	1.571**	.094
Missing Track in 10th	.323**	.088	.334**	.089	.327**	.089
Male			-.086	.069	-.092	.069
Black			-.146	.117	-.105	.125
Hispanic			-.139	.126	-.061	.131
Asian			-.430*	.191	-.377*	.192
Siblings			.042*	.021	.040	.021
Non-English Home			.087	.057	.055	.057
Gifted			-.196	.106	-.200	.106
Discuss			-.004	.049	.002	.049
Moved			.018	.091	.023	.091
Married			-.086	.138	-.083	.139
Divorced			.044	.127	.053	.127
SES			-.331**	.053	-.296**	.054
%Single Parents					.067	.058
% LEP					-.092*	.042
% FR Lunch					.007**	.002
% White					.005*	.002
% Parent Volunteer					-.003	.004
Magnet					.341*	.150
Catholic					-.545*	.271
Public Choice					-.268**	.099

[β] Marginally significant with p=.055.
[γ] Marginally significant with p=.067.

As we saw in the early elementary grades, prior student achievement was always a strong predictor of students' group assignments. However, even conditional on prior achievement, learning behaviors were important as well, at least for admission to the academic track. Both student and teacher reports were statistically and substantively significant across models in tables 14.1 and 14.2. I also ran models similar to Model 1 but omitting measures of learning behaviors, and found coefficients for reading and math achievement were only slightly higher than those reported in Model 1. This indicates that, while learning behaviors were an important predictor of student achievement, their inclusion in the models predicting track assignments provided additional, unique information not reflected in students' prior achievement. The coefficients for teacher and student reports of learning behaviors are not directly comparable, as they have different ranges and variances. However, teacher reports have larger standardized effect sizes and are more consistently significant across models. Based on the estimates in Model 3 in table 13.1, a one SD increase in the 8th grade student self-reported index of learning behaviors was associated with a .125 increase in the log odds of membership in academic track (versus the general track) in 10th grade, and a similar one SD increase in the teacher report of learning behaviors was associated with a .209 increase. By comparison, one SD increases in the 8th grade math and reading test scores were associated with .312 and .284 increases in the log odds of academic track membership, respectively. Translated into probability, a typical student (i.e. a student with all covariates values equal to the grand means – the intercept in the model) had a mean estimated .373 probability of academic track assignment. *Ceteris paribus*, a student with self-reported learning behaviors one SD above the mean would have seen this estimated probability increase by about two and a half percentage points (to .398) and a student with teacher-reported learning behaviors one SD above the mean would have expected an increase of

about four percentage points (to .415). The measures of teacher- and student-reported learning behaviors were correlated, but only on the order of about .3, so it is not unreasonable to consider them simultaneously in the model. Only teacher reports of learning behaviors predicted vocational group membership in 10th grade, and neither predicted 12th grade vocational track membership.

As in the prior chapter, we see some "stickiness" to group assignments over time. Students in the highest ability group in 8th grade were more likely to be in the college-prep/academic track in high school, even conditional on achievement, learning behaviors, and other covariates. A number of student background characteristics also influenced track assignments net of students' academic and behavioral characteristics. Males were more likely to be in the vocational track in 10th grade, although not at the end of high school. Although African American students were less likely to be in the academic track on average, once controls were introduced they appeared to have an advantage in academic track assignment. This finding is unexpected given the large raw difference in track assignments favoring white students, but is not without precedent in the literature (Alexander, Cook, and McDill, 1978; Gamoran and Mare, 1989; Garet and DeLany, 1988; Jones et al., 1995). I consider this in greater detail in the next subsection. Time-varying socioeconomic status composites are significant covariates in all models they are included, positively predicting academic track membership and negatively predicting vocational track membership. Students with a greater number of siblings also had lower odds of academic track membership. Family disruptions such as marriages, divorces, and residential moves are generally not significant. Among the various measures of family involvement with children and their education (most of which were not included in the final models above but are described in Appendix 3), the only consistent predictor was how often

students reported discussing what they learned at school with their parents. Frequency of these discussions was positively associated with academic track membership in both waves.

Some school characteristics also influenced students track locations, but most of these effects were comparatively small and inconsistent across models. The conventional measure of school affluence, the proportion of the student body that qualifies for federally-subsidized school lunch, was generally not a strong predictor. Nor were the measures of overall school achievement (operationalized as the proportions of students in remedial classes)[4]. Students attending public schools of choice were less likely to be assigned to academic tracks. There was also suggestive evidence (in models not reported above) that students attending schools in less urban areas were more likely to be assigned to the academic tracks and less likely to be assigned to vocational tracks. These effects were generally explained by the inclusion of other school-level variables. Finally, by the end of high school, there was a substantial shift towards the academic/college-prep program and away from vocational programs in Catholics schools. The tendency of Catholic schools to use less tracking and push more students into the academic track is a well-known phenomenon in the literature (Bryk, Lee, and Holland, 1993; Coleman, Hoffer, and Kilgore, 1982; Gamoran and Mare, 1982).

A Closer Look at Race

One of the most persistent and pernicious findings in the research literature on grouping is that, whenever students are grouped according to achievement or perceived ability, racial minority students (particularly African Americans) are disproportionately assigned to lower groups. This is true in the NELS dataset as well. About 47% of white students reported being in the academic track at the end of high school, versus 43% of black students and 35% of Latinos.

[4] Not included in the models above.

Asian students were more likely to report being in the academic track (64%). However, the analyses in the preceding section suggest that, once statistical controls for other student and school characteristics are introduced, it appears that African American students were *more* likely to be assigned to the academic track throughout high school. One possible explanation is that black and white students might be segregated into different types of schools. Black students may be more likely to be assigned to higher tracks in schools with greater concentration of minority students, if they encounter less discrimination or are likely to choose to take advanced courses with some assurance that they would not be the only student of color in the class. In the analysis of within-group placements in primary school, we saw a similar advantage for black students explained by school and classroom composition. Below we consider the interaction between student race and school racial composition and their influences on individual student track assignments. Model 1 includes only a (random) intercept and dummy variables for race. Race variables are uncentered, so the reference group is white students. Model 2 adds controls for prior reading and math scores (grand-mean centered). Model 3 adds school racial composition (operationalized as percent white students, grand-mean centered) at level two to the intercept and race variables. Specifically:

Level-1 Model

$\text{Prob}[ACADEMIC = 1|\beta_j] = \phi_{1ij}$
$\text{Prob}[VOCATIONAL = 1|\beta_j] = \phi_{2ij}$
$\text{Prob}[GENERAL = 1|\beta_j] = \phi_{3ij} = 1 - \phi_{1ij} - \phi_{2ij}$
$\log[\phi_{1ij}/\phi_{3ij}] = \beta_{0j(1)} + \beta_{1j(1)}*(MATH1_{ij}) + \beta_{2j(1)}*(READ1_{ij}) + \beta_{3j(1)}*(DBLACK_{ij}) + \beta_{4j(1)}*(DHISPAN_{ij}) + \beta_{5j(1)}*(DASIAN_{ij})$
$\log[\phi_{2ij}/\phi_{3ij}] = \beta_{0j(2)} + \beta_{1j(2)}*(MATH1_{ij}) + \beta_{2j(2)}*(READ1_{ij}) + \beta_{3j(2)}*(DBLACK_{ij}) + \beta_{4j(2)}*(DHISPAN_{ij}) + \beta_{5j(2)}*(DASIAN_{ij})$

Level-2 Model

$\beta_{0(1)} = \gamma_{00(1)} + \gamma_{01(1)}*(SPWHITE2_j) + u_{0j(1)}$
$\beta_{1(1)} = \gamma_{10(1)}$
$\beta_{2(1)} = \gamma_{20(1)}$
$\beta_{3(1)} = \gamma_{30(1)} + \gamma_{31(1)}*(SPWHITE2_j)$
$\beta_{4(1)} = \gamma_{40(1)} + \gamma_{41(1)}*(SPWHITE2_j)$
$\beta_{5(1)} = \gamma_{50(1)} + \gamma_{51(1)}*(SPWHITE2_j)$

$\beta_{0(2)} = \gamma_{00(2)} + \gamma_{01(2)}*(SPWHITE2_j) + u_{0j(2)}$
$\beta_{1(2)} = \gamma_{10(2)}$
$\beta_{2(2)} = \gamma_{20(2)}$
$\beta_{3(2)} = \gamma_{30(2)} + \gamma_{31(2)}*(SPWHITE2_j)$
$\beta_{4(2)} = \gamma_{40(2)} + \gamma_{41(2)}*(SPWHITE2_j)$
$\beta_{5(2)} = \gamma_{50(2)} + \gamma_{51(2)}*(SPWHITE2_j)$

Results are as follows:

Table 16.1 – Student Race and Academic Track Assignment in 10th Grade

	Model 1		Model 2		Model 3	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Intercept γ_{00}	-.144**	.042	-.434**	.047	-.386**	.053
School %White γ_{01}					-.006**	.002
Prior Reading γ_{10}			.066**	.004	.067**	.004
Prior Math γ_{20}			.044**	.004	.044**	.004
Black γ_{30}	.025	.097	.592**	.106	.432**	.121
School %White γ_{31}					-.002	.004
Hispanic γ_{40}	-.223*	.088	.160	.095	.097	.111
School %White γ_{41}					.003	.003
Asian γ_{50}	.456**	.099	.332**	.106	.260*	.109
School %White γ_{51}					-.003	.004

Table 16.2 – Student Race and Academic Track Assignment in 12th Grade

	Model 1		Model 2		Model 3	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Intercept γ_{00}	.351**	.038	-.004	.042	.048	.049
School %White γ_{01}					-.004*	.002
Prior Reading γ_{10}			.081**	.004	.081**	.004
Prior Math γ_{20}			.041**	.004	.042**	.004
Black γ_{30}	-.185*	.087	.549**	.097	.540**	.112
School %White γ_{31}					.005	.004
Hispanic γ_{40}	-.492**	.078	.031	.085	.010	.098
School %White γ_{41}					.004	.003
Asian γ_{50}	.448**	.092	.306**	.102	.237*	.105
School %White γ_{51}					-.002	.004

Table 17.1 – Student Race and Vocational Track Assignment in 10th Grade

	Model 1		Model 2		Model 3	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Intercept γ_{00}	-1.704**	.051	-1.827**	.054	-1.794**	.062
School %White γ_{01}					-.003	.002
Prior Reading γ_{10}			-.040**	.006	-.039**	.006
Prior Math γ_{20}			-.036**	.006	-.036**	.006
Black γ_{30}	1.045**	.111	.685**	.115	.667**	.136
School %White γ_{31}					.003	.004
Hispanic γ_{40}	.494**	.106	.235*	.109	.071	.139
School %White γ_{41}					-.003	.004
Asian γ_{50}	.095	.163	.149	.168	.138	.172
School %White γ_{51}					.004	.006

Table 17.2 – Student Race and Vocational Track Assignment in 12th Grade

	Model 1		Model 2		Model 3	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Intercept γ_{00}	-1.148	.049	-1.201**	.050	-1.227**	.061
School %White γ_{01}					.002	.002
Prior Reading γ_{10}			-.017**	.005	-.017**	.005
Prior Math γ_{20}			-.017**	.005	-.017**	.005
Black γ_{30}	.030	.109	.113	.112	.145	.129
School %White γ_{31}					-.002	.004
Hispanic γ_{40}	.066	.097	-.067	.099	-.003	.115
School %White γ_{41}					.000	.004
Asian γ_{50}	-.426	.168	-.446*	.172	-.403*	.176
School %White γ_{51}					.000	.006

In tables 16.1 and 16.2, detailing the odds of assignment to the academic track, we can see the substantial unadjusted differences across racial groups in Model 1. Hispanic students were less likely to be assigned to the highest track, while Asian students were more likely. Black students had roughly similar academic track assignments in 10th grade but were slightly less likely to be in the academic track in 12th grade. But after controlling for prior achievement in Model 2, the adjusted odds of track assignment for Hispanic students was similar to their white majority peers, the advantage enjoyed by Asian students diminished, and African American students had greater odds of being assigned to the academic program. Variants of Model 2 with random slopes indicated that variance in school race slopes was not significant, but reliability was also low. Once controls for school composition are added in Model 3, the story remains largely the same. The mean advantages for black and Asian students change slightly, but remain statistically and substantively significant. School composition only significantly changes academic track assignment odds for white students, which decline as the proportion of white students in the school decreases.

As to vocational program membership, black students were substantially more likely to be in vocational programs in 10th grade, but this disparity diminished by 12th grade. This was probably mostly due to a general departure from the vocational track during the second half of high school for all students, as substantial numbers of vocational track students either move into general or academic programs, or else drop out of school. This is discussed in more detail in the next section. Adjusting for prior achievement reduces the black-white gap in vocational track enrollment in 10th grade and eliminates it in 12th grade. Adjusting for prior achievement also eliminates the white-Hispanic gap in both grades. Asian students were less likely to be in the vocational track by the end of high school net of all controls in Table 15.2 and adjusting for

school racial composition in Table 17.2 above. None of the estimates for school racial composition variables are significant. Overall, it appears that individual factors drove much of the selection into different tracks, and observed school characteristics provided comparatively less explanatory power.

Mobility Between Tracks

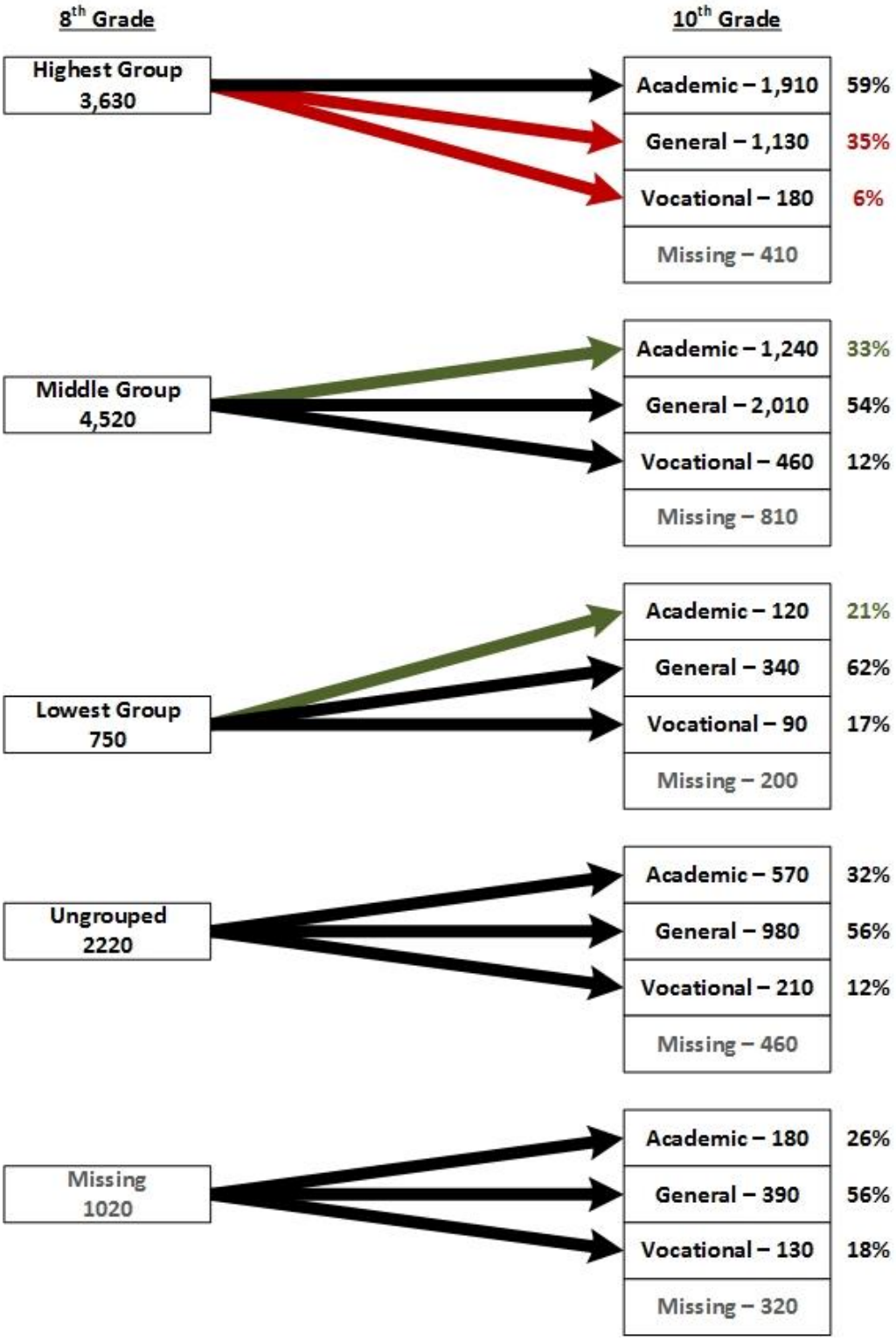
Mobility between group levels in high school

A common critique of tracking is that it makes it difficult or impossible for students assigned to lower tracks to move up once students are assigned to lower tracks. Slower learning progression in low track classes can make it hard to catch up with peers in more advanced classes. The stigma of being labeled a low performer can depress a student's academic self-concepts. Requirements of prerequisite courses and teacher or counselor referrals can act as barriers to entry (Mickelson and Everett, 2008). Rosembaum (1976) described progression through school as a tournament, where students in upper tracks could "lose" and move downward, but upward mobility was impossible. Interviews and observations conducted by Oakes (2005) and Page (1991) suggested upward mobility was rare. "To enter Southmoor as an academically unsuccessful student," Page explains, "is not simply to be different from the majority of students but to be irremediably different, and teachers are not held accountable for students' instruction." However, not all researchers have found the situation to be so dire. Metz (1978) observed that some "black children, many from poorer parts of the city... did their academic work well enough to qualify for Track Two [the second-highest track], some moving out of lower tracks as they caught on to required skills and behavior. And they were sufficiently diligent and cooperative to remain in Track Two or move up." Gamoran (1992), using data from

High School and Beyond, found substantial variation in track mobility between schools. He also found that track immobility led to lower overall achievement in mathematics and increased inequality in achievement between tracks. Hallinan (1996) found 30% of students changed tracks in English and 11% changed tracks in math during high school, and upward mobility was more common than downward. Subsequent work has generally found mobility is possible, although most students tend to be stable over time in secondary schools (Lucas, 1999; Lucas, 2001; Lucas and Good, 2001; McFarland and Rodan, 2009).

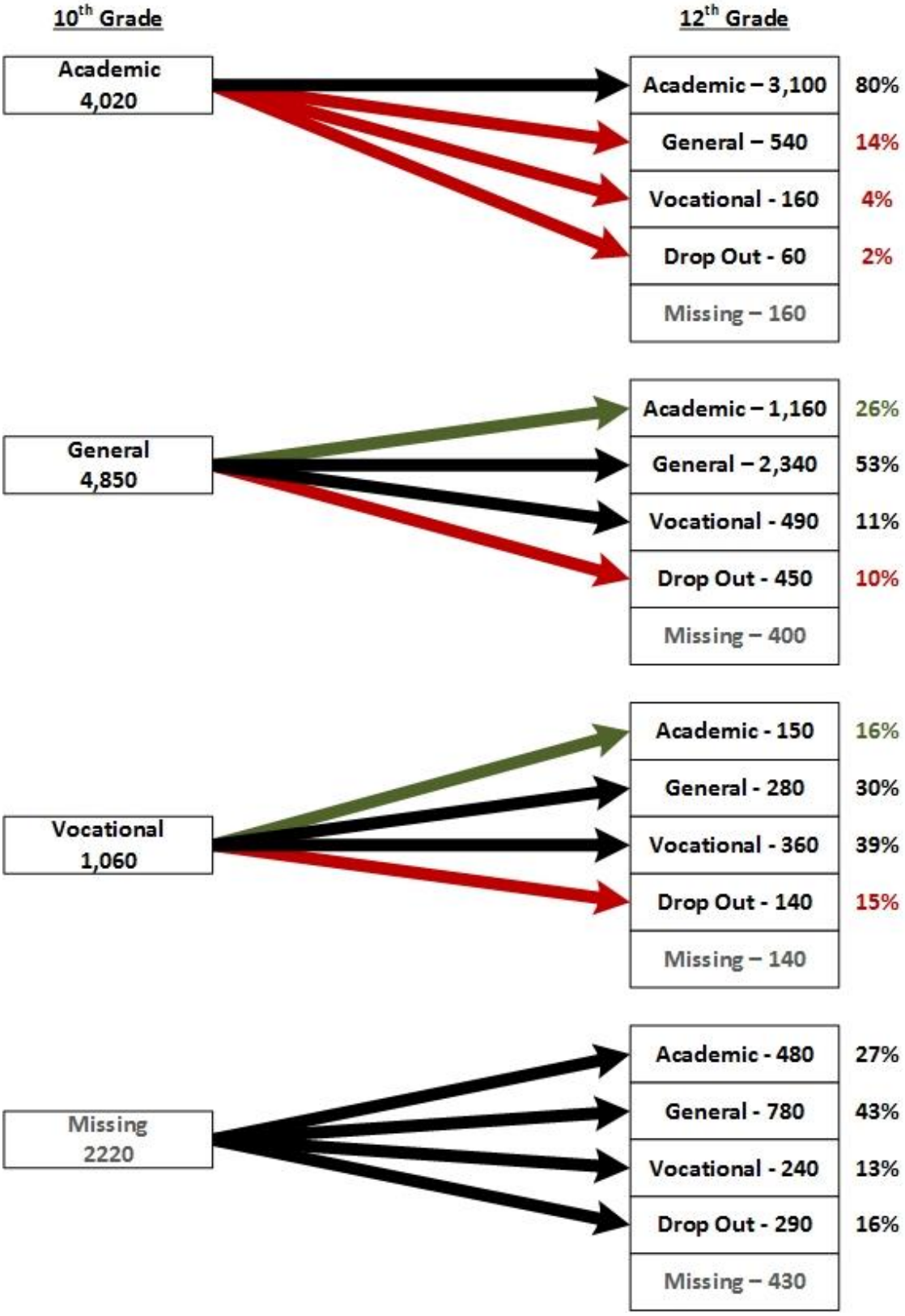
Two figures below describe mobility as students progress from 8th grade through the end of high school. An important caveat to the arrangement of tracks in these figures is that I am not satisfied that the vocational track should be considered strictly lower than the other two programs in a hierarchical sense. Vocational programs tend to differ qualitatively in important ways from generalized high school programs and the career-focused instruction they offer may be more appropriate and provide better outcomes for some students. However, it is true that students the vocational tracks had lower achievement test scores than students in either academic or general tracks. For the purposes of operationalizing mobility, I focus on movement in and out of the academic track. In the figures below, upward and downward mobility is indicated by green and red arrows, respectively. Following Lucas (2001), I also consider dropping out an alternative location that students can transition into after 10th grade in Figure 8.

Figure 7 – Track Mobility between 8th and 10th grades



There was substantial mobility between ability groups in 8th grade and academic tracks in 10th grade. Roughly one third of students in the highest group in 8th grade were in the general track two years later, and conversely, roughly one third of students in the middle group in 8th grade had moved into the highest track. Most students in the lowest ability group moved into the general track, although movement into the academic track was not rare. Students who were not grouped in 8th grade had aggregate track placements strikingly similar to those of middle group students. There were also 2220 students that were missing information on track assignments in 10th grade and 1320 students (including 238 who dropped out) missing data on track assignments in 12th grade. It is interesting that the proportion of missing data on this variable decreased substantially in the later wave, as usually the opposite occurs with longitudinal datasets. It is possible that students might have had more trouble identifying their track in 10th grade than at the end of their high school careers.

Figure 8 – Track Mobility between 10th and 12th grades



By the end of high school, track membership settled somewhat and mobility decreased. More students moved into the academic track than out of it. Also note that, if one does consider the general track to be higher than the vocational, upward mobility was predominant for these students as well. Slightly more than half of students in the vocational track in 10th grade had moved into the general or even academic tracks by the end of high school. On the other hand, students in the vocational track early in high school also had the highest dropout rates.

Predictors of group mobility

I ran a series of hierarchical logistic regressions to explore what characteristics of students and schools predicted mobility between group levels. Table 18.1 describes predictors of upward mobility between 8th and 10th grade. Students were coded as upwardly mobile if they were not in the highest ability group in 8th grade and reported being in the academic/college-prep program in 10th grade. Table 18.2 describes predictors of upward mobility between 10th and 12th grades. Students were coded as upwardly mobile if they reported that they were not in the academic track in 10th grade but were in 12th. In both sets of analyses, the sample was restricted to those students for whom upward mobility is possible, i.e. students that were in the highest groups in the prior wave were excluded. As with the analyses presented earlier in tables 4.1 through 5.2, these are two-level HLM models with students nested in the school they were in during the later wave. All school-level variables modify the (random) intercept and all slopes are fixed. All variables are centered around their grand means.

Table 18.1 – Predictors of Upward Mobility into the Academic Track between 8th and 10th grades

	Model 1		Model 2		Model 3	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Intercept	-1.174**	.053	-1.233**	.055	-1.220**	.056
LoGroup in 8th	-.279*	.128	-.255	.134	-.272*	.134
Ungrouped in 8th	-.456**	.150	-.399**	.154	-.396*	.153
Prior LB – S Report	.265**	.075	.251**	.078	.251**	.079
Prior LB – T Report	1.101**	.163	.966**	.168	.973**	.170
Prior Reading	.049**	.006	.041**	.006	.041**	.006
Prior Math	.060**	.006	.048**	.006	.048**	.006
Male			.040	.069	.031	.069
Black			.584**	.135	.546**	.143
Hispanic			.226	.138	.161	.148
Asian			.350*	.169	.317	.170
Siblings			-.082**	.024	-.081**	.024
Gifted			.502**	.113	.492**	.114
Home Language			-.097	.057	-.079	.057
SES			.481**	.054	.450**	.056
Prior Retained			-.380**	.111	-.382	.111
Discuss			.109*	.054	.108*	.054
Moved			-.067	.100	-.071	.100
Married			-.111	.165	-.109	.166
Divorced			-.073	.146	-.067	.146
%Single Parents					-.036	.088
% LEP					.016	.054
% FR Lunch					-.001	.003
% White					-.003	.002
% Parent Volunteer					.008*	.003
Magnet					-.072	.243
Catholic					-.166	.266
Public Choice					-.593**	.152

Table 18.2 – Predictors of Upward Mobility into the Academic Track between 10th and 12th grades

	Model 1		Model 2		Model 3	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Intercept	-1.277**	.047	-1.383**	.054	-1.368**	.055
Vocational in 10th	-.024	.107	.039	.116	.044	.117
Prior LB – S Report	.110	.080	.117	.090	.126	.090
Prior LB – T Report	.515**	.076	.501**	.090	.493**	.090
Prior Reading	.029**	.005	.023**	.006	.023**	.006
Prior Math	.060**	.006	.055**	.007	.056**	.007
Male			-.065	.084	-.063	.084
Black			.638**	.154	.563**	.160
Hispanic			.250	.138	.239	.141
Asian			.318	.163	.350*	.166
Siblings			-.081**	.027	-.080**	.027
Gifted			.175	.100	.181	.102
Home Language			-.039	.065	-.035	.067
Discuss			.133*	.063	.135*	.063
Moved			.105	.109	.114	.112
Married			-.091	.180	-.068	.181
Divorced			.126	.156	.098	.158
SES			.429**	.060	.432**	.062
%Single Parents					.028	.059
% LEP					-.092*	.043
% FR Lunch					.002	.003
% White					-.003	.002
Magnet					-.218	.181
Public Choice					-.183	.106
Catholic					.501**	.174
Prop Parent Volunteer					-.086	.057

The findings for mobility are convergent with those in tables 14.1 and 14.2 describing academic track assignments in general. Students with better prior achievement were more likely to move up, but students with better reported learning behaviors were also more likely to move up conditional on their achievement. Teacher reports of learning behaviors were a better predictor of mobility than student self-reports. Interestingly, students who were not assigned to ability groups for math in 8th grade were less likely to move up into the academic tracks, despite

having test scores comparable to students that were assigned to middle-level ability groups in 8th grade (the reference group in table 18.1). The effects of student background characteristics are familiar. Students with more siblings and those that were retained in 8th grade were less likely to be upwardly mobile. Gifted students, students from higher-SES families, and those who reported that they discussed what they've learned at school with their parents more frequently were more likely to move up. African American students (conditional on other covariates) were more likely to move up during both transition periods. School-level covariates generally had small effects, excepting school sector. Students in public schools of choice experienced less upward mobility early in high school (conditional on covariates), and students in Catholic schools were much more likely to move into the academic track by the end of high school.

Effects of Group Assignments

Academic outcomes

By the time students enter high school, they have generally experienced both grouped and ungrouped instruction in different contexts for many years. A large majority of the students in the sample reported being grouped by ability in 8th grade. In the previous chapter, I found that within-class grouping practices that a majority of students encounter in the early primary grades can contribute to increased inequality in both academic outcomes and learning behaviors. It is possible that sorting students into tracks in high school might compound this divergence, leading to even larger effects for group assignments. On the other hand, it is also possible that by the time students enter high school, they have settled into regular patterns of achievement and behaviors, and be less sensitive to the different instructional climates offered in different instructional programs. Carbonaro (2005) found substantial between-track disparities in student

academic effort, but also that these differences were largely the result of differences that existed before students entered high school.

Separate propensity scores were estimated for academic and vocational track assignments in 10th and 12th grade, respectively. For each analysis the sample was divided into strata and balanced on treatment and control cases on pre-treatment covariates from the prior wave and any time-varying covariates collected during the wave that could not have been caused by treatment (e.g. family SES). The region of common support lay between probabilities of treatment of .02 and .91 for academic track assignment in 10th grade and .02 and .98 in 12th grade, but only between .01 and .70 for vocational track assignments in 10th grade and .01 to .78 in 12th grade. The final analytic model was a two-level HLM. The first level included the treatment, a vector of dummy variables for propensity strata, and the natural log of the propensity score. The intercept and the slope for the treatment were allowed to vary. The tables below report the point estimate and standard error for the effect of track membership, and the standard deviation of the school mean of track slopes (square root of the variance τ_{11}). I also calculate the standardized effect size for continuous outcomes and the mean difference in probabilities between treated and untreated students for dichotomous outcomes to give an indication of the relative strength of the effects.

Table 19.1 – Track Effects on Academic Outcomes

	Reading				Math			
Treatment	**Coefficient**	**Standard Error**	**Standardized Effect Size**	**Std. Dev. of Track Slope**	**Coefficient**	**Standard Error**	**Standardized Effect Size**	**Std. Dev. of Track Slope**
Academic Track in 10th Grade	1.003**	.186	.099	1.12	1.026**	.159	.104	1.18*
Academic Track in 12th Grade	.362	.251	.034	1.28	.941**	.211	.090	.681
Vocational Track in 10th Grade	-1.261**	.265	-.124	.591	-.934**	.255	-.091	1.66*
Vocational Track in 12th Grade	-.311	.374	-.029	1.24	-.904**	.345	-.086	2.07

Table 19.2 – Track Effects on Learning Behaviors

	Student Report				Teacher Report			
Treatment	**Coefficient**	**Standard Error**	**Standardized Effect Size**	**Std. Dev. of Track Slope**	**Coefficient**	**Standard Error**	**Standardized Effect Size**	**Std. Dev. of Track Slope**
Academic Track in 10th Grade	.025*	.011	.053	.038	.050**	.013	.086	.132*
Academic Track in 12th Grade	.070**	.018	.117	.095**	.024	.019	.038	.125*
Vocational Track in 10th Grade	-.010	.018	-.021	.057	-.030	.022	-.052	.159
Vocational Track in 12th Grade	.020	.024	.033	.121**	.011	.025	.017	.022

Conditional on the assumptions of the model, track membership altered both academic outcomes and learning behaviors. Assignment to the academic track improved reading and math test scores in both waves. However, the effect sizes were substantively small, and decreased between 10th and 12th grades. It also improved the student reports of learning behaviors by similarly-small amounts. Teacher reports of learning behaviors improved for academic track sophomores but not seniors. In contrast, students in vocational programs saw depressed math and reading outcomes. However, effects on learning behaviors were generally weak, with only small decreases in teacher reports of learning behaviors in the 10th grade reaching statistical significance at conventional levels. There is some school-level variation in the estimates, statistically significant in a few of the models for math achievement scores and teacher reports of learning behavior. Analysis of stratum-by-treatment interaction effects (not reported here) indicated that treatment effects were generally consistent across strata. I also estimated the log odds of dropping out by the end of the second wave of data as an outcome for track membership in 10th grade. Membership in the academic program in 10th grade was associated with a population-average school mean decrease in the probability of dropping out from 3.4 percent to 1.3 percent. Despite the substantial raw mean differences in dropout rates between students enrolled the vocational and general tracks in 10th grade (recall Figure 8), vocational track membership was not associated with lower drop out after adjustment for propensity scores.

The estimates reported above represent causal estimates under two key assumptions: strong ignorability and the SUTVA. The strong ignorability assumption – that group assignments are independent of the potential outcomes conditional on the observed covariates in the propensity score model – cannot be empirically verified. However, sensitivity analysis can

assess how robust the model estimates are to potential unobserved confounding variables. In tables 14.1 through 15.2, we found that student family socioeconomic status was generally the strongest predictor of treatment (track) assignments other than prior academic outcomes. I therefore use student SES as the basis for estimating the potential bias introduced by a hypothetical unobserved confounding variable. This potential bias is calculated by multiplying the partial regression coefficient from the propensity score model (estimating the outcome of interest) by the mean difference in student SES between treated and control students. All statistically significant results from tables 19.1 and 19.2 are robust to such a potential unobserved confounding variable. For example, for academic track assignments in tenth grade, there is a potential bias of .051 (versus a coefficient of 1.003) for reading achievement, .111 (versus a coefficient of 1.026) for math achievement, and .002 (versus a coefficient of .025) for student-reported learning behaviors. For vocational track assignments in 10th grade, potential bias is .080 (versus a coefficient of -1.261) for reading achievement, .218 (versus a coefficient of -.934) for math, and .001 (versus a coefficient of -.030) for teacher-reported learning behaviors. It is important to remember, however, that these effects could vary across schools by amounts that make it reasonable to suggest that effect sizes are substantially higher in some schools but close to zero in others[5]. The Stable Unit Treatment Value Assumption – which stipulates that there is only one version of the treatment and that the treatment assignment of one student does not affect the potential outcomes of others – is more problematic. Both components of this assumption are likely violated in real educational circumstances. The semi-organized curricular programs I refer to as "tracks" in this chapter differ substantially from school to school, especially among vocational programs that focus on different fields of study. Students can be influenced by the treatment assignments of others in a variety of ways, such stigmatization when a student is

[5] Note the standard deviations of track effect sizes reported in the tables above.

assigned to a lower track than his peers or reluctance of minority students to enroll in upper-track courses dominated by white students. It is unclear how violation of this assumption might bias my results.

Labor Market Outcomes

As discussed in in the introductory chapter, schools not only communicate academic knowledge to students, but serve as training grounds to prepare students to assume adult roles. Many of the learning behaviors I have focused on in this chapter (consistent attendance and punctuality, attention to instructions, assignment completion, arriving prepared with necessary materials) translate into important behaviors in the workplace. If students' experiences in grouped instruction shape the development of their learning behaviors, they may also shape students' future success in the labor market. To test this, I examine several outcomes from data collected following students' secondary education careers. Follow-up surveys were administered in 1994 (two years after most students graduated) and 2000. From student reports of their post-secondary education experiences in the 1994 survey, I consider two outcomes: whether respondents reported being enrolled in a traditional college bachelor degree program, and whether or not they were enrolled in or had completed any other formal postsecondary degree program such as an associate's degree, certificate, or professional license. From the 2000 survey, I consider four outcomes: whether or not respondents reported having a full time job at the time the survey was administered, the number of weeks they worked in the prior year, hourly wage rates for their current or most recent job, and annual personal income the prior year.

Table 20 – Track Effects on Postsecondary Education Enrollment in 1994

	College Enrollment				Other Postsecondary Education			
Treatment	**Coefficient**	**Standard Error**	**Probability Difference**	**Std. Dev. of Track Slope**	**Coefficient**	**Standard Error**	**Probability Difference**	**Std. Dev. of Track Slope**
Academic Track in 12th Grade	.704**	.069	.342 / .498	.199	-.290**	.071	.335 / .283	.665
Vocational Track in 12th Grade	-.662**	.136	.207 / .120	.059	.169*	.080	.396 / .435	.099

Table 21 – Track Effects Labor Market Outcomes in 2000

	Has Full Time Job				Hourly Wage[δ]			
Treatment	**Coefficient**	**Standard Error**	**Probability Difference**	**Std. Dev. of Track Slope**	**Coefficient**	**Standard Error**	**Standardized Effect Size**	**Std. Dev. of Track Slope**
Academic Track in 12th Grade	-.029	.086	.792 / .787	.067	.834*	.337	.076	2.05
Vocational Track in 12th Grade	.197	.127	.802 / .831	.017	.362	.425	.033	1.83

	Weeks Worked				Income			
Treatment	**Coefficient**	**Standard Error**	**Standardized Effect Size**	**Std. Dev. of Track Slope**	**Coefficient**	**Standard Error**	**Standardized Effect Size**	**Std. Dev. of Track Slope**
Academic Track in 12th Grade	-.271	.332	-.025	.652	995	605	.049	2169
Vocational Track in 12th Grade	.587	.437	.053	.575	85.3	834	.004	4876

[δ] Includes both employed and unemployed individuals. Unemployed respondents provided the wage of their most recent job.

Conditional on the assumptions of the model, track assignments appear to have significantly affected students' post-secondary education trajectories. Students assigned to the college-preparatory track were indeed substantially more likely to enroll both in traditional bachelor degree programs. Students in the vocational tracks were less likely to enroll in 4-year college programs. It is possible, however, that this need not be considered a negative outcome. Vocational tracks are designed to provide students with skills valued in the labor market, so it is possible that graduates of these programs are choosing to put those skills to use immediately after high school. Students in vocational tracks were also more likely to pursue postsecondary education options other than traditional 4-year college degrees.

However, the findings in Table 21 provide little evidence for track effects on labor market outcomes by the final wave of data collection. Students that graduated from the academic track earned an average of about 83 cents more per hour eight years after high school. Vocational track membership does not appear to have changed wage rates. Neither program seems to have affected the number of weeks worked in the prior year. There is slight, suggestive evidence that the workplace-oriented training provided by the vocational tracks provided benefits to students as they moved into the labor market. Graduates of vocational programs were about 3 percentage points more likely to hold a full time job in 2000 than general track graduates. This difference was not quite significant in the unit-specific model, but marginally significant ($p = .050$) as a population average, reflecting a more precise estimate and smaller standard error.

It is actually quite interesting that, despite a general perception that the college preparatory tracks offers the highest-quality instruction and greatest opportunities for students, there was not a larger disparity in labor market outcomes. Wages were higher for academic track graduates (compared to similar general track graduates), but the gap is less than one dollar per

hour. The overall disparity in annual income was less than one thousand dollars, an only marginally-significant difference (p=.10). This is especially surprising given the substantial differences in the pursuit of different types of post-secondary education. However, it that the measurement of wages eight years after high school graduation might not adequately capture long-term differences in earning potential. Students that from graduated four-year college programs have had less time on the labor market, but have likely acquired skills that will provided greater growth in earnings over the course of their careers.

Discussion and Conclusion

Returning the to the central motivating question of this research – Can students work their way up into higher groups as they progress through school? – the answer in secondary schools during this study period is a qualified "Yes." There were considerable opportunities to access the college-preparatory track, and students could improve their chances of upward mobility by demonstrating good learning behaviors and raising their achievement. A majority of students reported that their group assignments were largely a result of their own choices, usually made in consultation with others. However, choice was not equally available to all students. Moreover, as we saw in primary schools, group assignments could compound existing differences in achievement and behaviors. But by the time students are moving through secondary schools, all three of these factors – group assignments, achievement, and learning behaviors – are more stable than earlier in their educational careers. The estimated causal effects of group assignments were relatively small, generally on the order of .1 SD. By the end of high school, program assignment effects on learning behaviors are practically negligible. This is

consistent with Carbonaro (2005) who, using data from the first two waves of the NELS dataset, found measures of student effort were largely settled before students entered high school.

The preponderance of evidence presented in this chapter stands against one critique of between-class grouping in high schools – that once students are relegated to lower-tracks, it is difficult or impossible to escape. It also suggests a more nuanced view of a second critique – that tracking systems tend to exclude minority students from the highest tracks. Numerous authors have provided detailed evidence of dramatic racial disparities in track assignments (Lewis and Diamond, 2015; Oaks, 2005; Page, 1991; Tyson, 2011). Across schools in the NELS dataset, black students were not significantly less like to report being enrolled in the academic track in 10th grade and only slightly less likely in 12th grade (43% versus 47.5% for white students). In terms of program assignment, my findings indicate that largest disadvantages were borne by Latino students, and the disparities were not explained by differences in reading achievement, home language environment, or immigration status.

Another important finding of this chapter is that, when one compares students with similar achievement profiles (a tenuous proposition when large aggregate inequalities exist between groups), African American students were more likely to be assigned to the highest track. This finding was not explained by school segregation. Clearly, there was something unobserved that improved the odds of high-track assignment for black students. The reasons for this apparent advantage are difficult to discern with the available data. It could be that these students were more able, or demonstrated promise in ways that were not reflected in the available covariates. Another speculative explanation is that by the late 1980s and early 1990s, teachers and other school personnel were well aware of the disparities in track assignments between black and whites students, and might have taken particular care to make more demanding coursework

available to minority students that could do well in them. Yet another possibility could be greater motivation among black students to attend college as a means of upward social and economic mobility. Research evaluating oppositional culture theories in education has sometimes found greater pro-school attitudes and desire to continue education among black students than whites (Ainsworth-Darnell and Downey, 1998; Ferguson, 2007; Lewis and Diamond, 2015). Regardless of the mechanism, this advantage in track assignment is important because it runs counter to traditional argument of structural bias against African American students. While detracking has been suggested as a key intervention for reducing black-white test achievement gaps (with some implementations seeing success, see Burris 2014 for reviews and discussion), it has also been pointed out that tracking can benefit talented minority students the most. While white students are more likely to come from middle-class or affluent families that can provide academic enrichment and challenge outside of school, minority students are more likely to be reliant on public schools for access to advanced learning opportunities (Loveless, 2016; Rosenbaum, 1999).

It should be noted that, while a conditional advantage in track assignments for African American students has been documented in the literature before (Alexander, Cook, and McDill, 1978; Gamoran and Mare, 1989; Garet and DeLany, 1988; Jones et al., 1995), this finding has not been validated by other research using the NELS dataset. The divergence seems to be driven (at least in part) by differences in how track memberships are defined. Argys, Rees, and Brewer (1996), in response to concerns over the reliability of student self-reports of track membership, focused on teacher reports of class descriptions to define student track positions. They found advantages in high track assignments for Hispanic students but not other groups. Sakura-Lemessy, Carter-Tellison, and Sakura-Lemessy (2009) operationalized tracks as a three-category

variable based on the number of course units students completed in certain subjects, and also restricted their sample to students that did not go on to attend college. They found that African American students were more likely to participate in "integrated" and vocational curricula versus an academic curriculum. It is possible that the students' reports of their program assignments did not reflect the reality of their structural locations. I chose to focus on self-reports of tracking for two reasons. First, because I was interested in group mobility, it was important to have a definition of track membership that was consistent and comparable across waves of data. Student reports were available for nearly all students, whereas teacher reports were not and could vary depending on which subject sampled teachers happened to teach. Using the number of courses taken by students works well for defining high school programs over four years, but there is much less variation in 10th grade.[6] Second, one of the ways track memberships are supposed to affect academic outcomes and learning behaviors is through social psychological mechanisms such as stigma attached to low group membership, reflected glory of high group membership, and Big Fish Little Pond effects. Defining group membership based on how students perceive themselves aligned my treatment definition with how students are likely to experience these factors.

The findings concerning labor market outcomes are interesting in the context of those regarding learning behaviors. While academic track membership substantially improved college enrollment, it does not improve later employment outcomes with the exception of a small increase in average wage rates. Perhaps this might be due to the similar development in learning behaviors between students across different programs by the end of high school? Moller and Stearns (2012), using the NELS dataset and defining four tracks based on transcript data, found

[6] I did attempt to use the student self-reports to create cross-validated course -based measures of track membership (i.e. define tracks by the typical coursework undertaken students that self-identified as members of different programs), but variation was substantial and I was unable to create convincing definitions in this manner.

that higher track membership was associated with higher incomes in the final wave of data collection in 2000. Most of the effect was mediated by college enrollment, but they also found that track membership improved annual earnings independent of the "quantity of education" students completed. They considered learning behaviors as well, but as control variables. My findings suggest that this might be undesirable, as learning behaviors can be caused by track membership (Hong and Hong, 2009 make a similar argument cautioning against adjustment for potential outcomes of grouping practices). However, the findings in the chapter are generally consistent with Moller and Stearns, with high-track membership associated with greater college attendance, improved hourly earnings, and (marginally significant) higher annual income.

Finally, it is important to note that, even if overall average effects of high school tracking polices are small on average, that there can be large, important effects in individual schools. The results presented in this chapter do not invalidate the observations of disturbing between-class disparities documented by Oakes (2005) and other critics of tracking. The adjustment for many prior covariates in the propensity score-based analysis, including prior group assignments, helps isolate the effects of individual treatment but can mask large raw differences in outcomes between students in different tracks. The key takeaway is that, while prior differences in achievement and learning behaviors explain much of the between-track inequality in these outcomes, the assignment of students to different high school programs can still exacerbate these differences. Track assignments may not be overtly biased against black students, but student background played a larger role in group assignments in secondary schools than they did in primary schools. Students from disadvantaged socioeconomic backgrounds were less likely access the higher tracks in every model considered. Track assignments can also have long-term

consequences for students, especially as they shape trajectories toward college, compromising social mobility.

Chapter 4 – Review and Concluding Thoughts

Context and Motivation

This project began as an attempt to reconcile two competing perspectives on the practice of ability grouping I confronted as an aspiring sociologist of education. One the one hand, my uniformly-positive personal experiences with grouping and my academic training to regard schools as organizations suggest that grouping is a rational, efficient means of organizing instruction. On the other hand, progressive educators and researchers generally oppose grouping on the grounds of social justice. The coherent narrative advanced by Ray Rist in *The Urban School* and Jeannie Oakes in *Keeping Track* – that grouping permanently handicaps poor and minority students with stigma and substandard instruction – remains a powerful influence in the education community. For example, at the time of this writing, if one enters "school tracking" into Google's search engine, one of the top results is the National Education Association's "Research Spotlight on Academic Ability Grouping." It explains, in part:

> Ability grouping, also known as tracking, is the practice of grouping children together according to their talents in the classroom. At the elementary school level, the divisions sound harmless enough - kids are divided into the Bluebirds and Redbirds. But in secondary schools, the stratification becomes more obvious as students assume their places in the tracking system.
>
> In many instances, these students are given labels that stay with them as they move from grade to grade. For those on the lower tracks, a steady diet of lower expectations leads to a low level of motivation toward school. Consequently, in high school, the groups formerly known as the Bluebirds and Redbirds have evolved into tracks: College Preparatory and Vocational.

The review provides a single peer-reviewed source (Cooper, 1996) and then offers the NEA's policy recommendation: "The National Education Association supports the elimination of such groupings. NEA believes that the use of discriminatory academic tracking based on economic status, ethnicity, race, or gender must be eliminated in all public school settings (NEA Resolutions B-16, 1998, 2005)." (NEA website, 2016). Both the technical-organizational

viewpoint and the narrative of oppression focus strongly on social structure and little on student agency – group assignments are something that are imposed on students. I wanted to understand to what extent students could shape their own fortunes within differentiated learning environments and how they were in turned shaped by their experiences. Are students doomed once they are relegated to low ability groups shortly after they enter kindergarten, or can they work their way up? As expected, my investigation did not provide a simple answer to this question. In this chapter, I review key findings from my research and how they fit in with existing literature, examine the limitations of the data and my methodology, and discuss possibilities for future research.

Key Findings

A surprising amount of mobility existed between group levels, particularly in the early primary grades.

Students in the lowest group in kindergarten had about a 50% chance of moving into a higher group by first grade, and students in the lowest group in first grade had a similar chance of moving up by third grade. Also, about a third of students in grouped classes in the ECLS-K changed groups during the year, and upward mobility was five times more common than downward. These within-year mobility rates are generally consistent with the upper range of estimates reported in other research (Dreeben, 1984; Eder, 1983; Groff, 1962; Hawkins, 1966; Rist,1970; Weinstein, 1976), and the findings concerning between-year mobility are the most comprehensive available to date. In the NELS dataset, mobility rates were somewhat lower by the time students enter high school, and declined as student near graduation. However, even after tenth grade, mobility was hardly impossible. More than a quarter of students in the general track

in 10th grade reported being in the academic/college-prep track by the end of high school, and more than half of the students in the vocational track in tenth grade moved into either the academic or general tracks by twelfth grade. Downward mobility out of the academic track was also possible, but less common than upward mobility into it. This level of mobility is higher than generally predicted by critics of tracking, and surprised my colleagues in the Tracking and Detracking Special Interest Group at an annual meeting of the American Educational Research Association. When Oakes and Guiton (1995) interviewed fifty teachers from several West coast high schools at around the same time the NELS dataset was collected, only sixteen could even recall a student that entered their school with "low skills" but improved enough to switch to the college-prep track from the general. My findings are more consistent with Hallinan (1996), who, using administrative data from two Midwestern cities, found "considerably more mobility than is typically assumed." She operationalized track assignments differently and separated math and English tracks, so the mobility rates she calculated were not directly comparable to my estimates in Chapter 3, but they are roughly similar in magnitude. She also found upward mobility was more common than downward, but (in contrast with my findings) found more mobility occurred later in students' high school careers than earlier.

The results concerning mobility rates do beg an important question regarding how much mobility should be considered desirable at each stage of education. On the one hand, mobility should be considered a positive factor, as instruction can be adapted to changes in student preparation, and prior errors in classification can be rectified. On the other hand, high rates of mobility suggest that either the rank order of student achievement is constantly shifting, or that group assignments are prone to high levels of error, potentially undermining the effectiveness of grouped instruction. It would be interesting to know if there are certain levels of mobility that are

optimal for student learning (Gamoran, 1992; Sørenson, 1970) , but such a question is beyond the scope of the current study.

Learning behaviors affect group placements and mobility.

My findings indicate that access to higher group assignments, including upward mobility following lower assignments, is predicted by students' prior learning behaviors, even controlling for prior achievement. This is true in both the early grades, where group assignments are made exclusively by teachers, and in high school, where group assignments are made on the basis of student choices, school policies, and the input of parents and school personnel. In this respect, it seems students really can work their way up. Or rather, at least some students can. There is strong evidence that mobility is more difficult for some students than others. In both datasets, I found that students from families with advantaged socioeconomic backgrounds had higher placement and mobility rates that their peers, even conditional on achievement and learning behaviors. While I did not find evidence of bias against African American students, Latino students were less likely to be in academic tracks in high school even controlling for academic and behavioral characteristics. Classroom and school context mattered, although not greatly. Attending an elementary class with greater proportion of low-achieving students made placing into a higher group easier, but attending a mostly white classroom made it more difficult. School-level factors were less influential in high schools, although school sector did have substantial effects on group placements and mobility.

But, group assignments do tend to stick with students.

Across nearly all models in both datasets, prior group assignments were also important predictors of later group assignments, even conditional on all available covariates. This could be due to the presence of student characteristics that are not observed in these survey-based datasets but known those making group assignments in schools. Another possibility is that the act of "labeling" students influences the future perceptions of these students by decision makers. As students age, these decision makers may even include the students themselves, as they internalize messages regarding their expected achievement and place within schools (Dreeben, 1968). This "stickiness" of group assignments was observed in Kerckhoff's research utilizing a long term panel study in Britain that followed a cohort of children from birth through adulthood. Students' nursery school group placements tended to stick with students into junior school, even controlling for test scores and teachers' ratings of students. Similarly, junior school group placements predicted the type of secondary schools students entered, which then affected secondary education credentials students obtained, and eventually students' success in the labor market. Kerckhoff describes this as "institutional inertia" in the placement of students within the structure of schools. "Once they were identified as having either high or low academic ability, that classification tended to stay with them," he explained. "It is apparent that some kind of labeling process was at work that, whatever the students' later performances, their earlier group placements continued to affect their placements in the structure." (Kerckhoff, 1993) While no similar long-term panel study of students that includes ability group placements exists in the United States, my findings from two separate panel studies of different periods of educational development are consistent with Kerckhoff's. Gamoran (1989) suggested that the group placements of students can become institutionalized, where "groups take on symbolic

characteristics that describe students' statuses independent of achievement." He found evidence of group placements sticking to students between first and second grade in several Chicago elementary schools. While "groups were not created through ritualistic assignment procedures... teachers appeared to carefully consider the distribution of prior achievement among students when assigning them to second-grade reading groups." He further speculated that "greater use of between-class grouping at higher grades might make group ranks become more salient as students move through the school system. Moreover, the longer students belong to a certain kind of group, the stronger their identification with that group may grow, from their own perspective and that of others." My findings indicate that, while mobility is certainly still possible, students do tend to settle into particular tracks in high school. Students that were in highest ability groups in eighth grade were more likely to have been in the highest track in 10th grade (net of prior achievement, learning behaviors, and all other covariates), and that track placements in 10th grade were powerful predictors of placements at the end of high school.

Group placements affect the development of both academic achievement and learning behaviors, but this influence diminishes over time.

By comparing groups of students with similar probabilities of treatment and balanced on large numbers of covariates, I found that assignment to higher-ranked groups was associated with higher student achievement in reading in the early primary grades and higher reading and math scores in high school. Attending the academic track (versus the general track) in tenth grade also substantially reduced the probability of dropping out of school. Conversely, lower group assignments predicted depressed academic outcomes. These findings are convergent with the substantial body of evidence for Matthew effects on academic outcomes (Hanushek and

Wößmann's, 2006; Lleras and Rangel, 2009; Oakes, 2005; Reuman, 1989; Tach and Farkas, 2006; Stanovich, 1986). Moreover, I found similar differential effects in the development of student learning behaviors. Conditional on the assumptions of the propensity score analysis, assignment to higher within-class reading groups in elementary school was associated with improvements in learning behaviors while students assigned to lower groups experienced less development in these behaviors. At the high school level, learning behaviors improved more for students enrolled in the academic tracks versus the general track, although there were no mean differences for vocational track students.

Several researchers have documented teachers justifying grouping practices on the basis of beliefs that students have different, relatively-fixed levels of ability, motivation, interest in schoolwork, and willingness to engage in the behaviors necessary to learn (Finley, 1984; Haller and Waterman, 1985; Oakes and Guiton, 1995). Oakes and Guiton (1995) contended that "schools saw their job as offering programs that *accommodate* rather than *alter* their students' ability and motivation" (emphasis in original). Because I control for differences in prior learning behaviors (among many other factors) in all of my models with learning behaviors as outcomes, my findings stand in sharp contrast with this belief. The disparities in learning behaviors observed between group levels can only partially be explained by selection. Rather, it is evident that something about students' experiences in grouped instruction is inspiring or dissuading them from participating more actively in the learning process.

Nonetheless, it does appear that the influence of group assignments diminishes as students age. Standardized effect sizes were on the order of .15-.20 in the early primary grades, but only about .10 or lower in high school. It is likely that, as students age, their identities and self-concepts as students become more stable, and they settle into particular patterns of behavior.

Based on his analysis of the NELS dataset, Carbonaro (2005) suggested that patterns of student effort directed towards learning was largely set before students entered high school, and thus less susceptible to the influences of track assignments. Although I found that the effects of group assignment were persistent even after several years in the early elementary grades, I was unable to determine whether these effects were cumulative as students moved through grades, potentially receiving different combinations of group placements along the way.[1] The preponderance of evidence suggests that interventions to improve the instruction and experiences of students assigned to low groups will likely be more effective if they are introduced in earlier grades.

In high schools, students report exercising considerable control over their group placements.

Student choice has been a neglected aspect of the literature on tracking at the high school level, and I was interested to discover the prominent role that students reported in the selection of their high school programs. Four of out five students that reported their track assignment in tenth grade indicated that they were involved in choosing which program to pursue. 58% of students reported that the choice was exclusively their own, while 22% reported that both that they were assigned to their track but also chose to be a part of it. To be sure, not all students faced the same options or constraints. Students with higher eighth-grade test scores or who were in high ability groups in the eighth grade were more likely to report that they had chosen their program. Asian students were more likely to report having chosen, while black and Hispanic students were less likely (than white students). Even within the same school, it was very common for some students to report that they could choose their track while others reported that they were assigned by the

[1] I attempted to use the inverse probability of treatment weighting methods described in Hong and Raudenbush (2008) to assess potential cumulative or interaction effects of repeated treatments (i.e. group assignments) over time, but I was unable to achieve balance across treatment groups for more than a single wave of treatments.

school. Membership in a vocational track was nearly always the result of deliberate student choice. I would speculate that perhaps this aspect of student volition might explain why assignment to the vocational track is associated with lower math and reading achievement outcomes (versus the general track), but differences in learning behaviors are negligible. It seems plausible that, while the training offered in vocational programs provides less-rigorous instruction in core academic subjects, the coursework offers practical training desired by self-selected students and helps keep them engaged in learning.

Group assignments in high school can have lasting consequences after students graduate.

One of my reasons for focusing on learning behaviors in this research is that, while academic knowledge is useful in many circumstances, the productive habits and routines of directed action developed in school are essential as students transition into adult roles. The results of propensity score analyses indicate that the college-preparatory tracks truly live up to their name. Compared to similar students in the general track, students enrolled in the academic track during their senior year were considerably more likely to attend a four-year college. Students enrolled in vocational tracks were less likely to attend four-year colleges, but more likely to pursue alternative avenues of postsecondary education such as certifications or associate's degrees. However, labor market outcomes were not radically different. Graduates of the academic track earned an average of 83 cents more per hour than graduates of the general track (after adjustment for propensity scores). The average employment rates and the number of weeks worked eight years after high school were no different across all former members of all three tracks.

Limitations of the Current Study

As with all scientific research, this project suffers from drawbacks that limit the extent to which we can draw conclusions for its results. Perhaps the most obvious limitation of the current research is that, while the ECLS-K and NELS datasets are excellent resources and the best available datasets to consider my research questions, both are rather dated. The last wave of data that included ability group information in the ECLS-K was conducted in 2002 and most of the students in the NELS graduated high school in 1992. There is an open question as to what extent the findings can be generalized to the experiences of students today. The period in which the NELS dataset was collected followed the "unremarked revolution" described by Lucas (1999) when traditional rigid tracking systems gave way to subject-specific grouping assignments as well as a time of state-level policy shifts away from the use of tracking at all (Loveless, 1999). Using data from the National Assessment of Education Progress, Loveless (2013) described trends in the use of ability grouping for reading and math instruction in fourth grade and tracking in eighth grade in more recent years. The findings indicate that, from 1992 to 2011, the use of ability grouping in fourth grade declined and then rebounded significantly. Tracking in 8th grade remained relatively steady during the same time period. Consider the tables below, all adapted from Loveless (2013):

Table 22 – Basis of Reading Instructional Groups (percent of students)

Year	Ability	Other	Not Created
2009	71	21	8
2007	64	20	15
2005	59	22	19
2003	47	25	27
2000	39	29	32
1998	28	33	39

Table 23 – Creation of Groups for Math Based on Ability (percent of students)

Year	Yes	No
2011	61	39
2009	54	46
2007	46	54
2006	49	51
2003	42	58
2000	41	59
1996	40	60
1992	48	52

Table 24 – Tracking in 8th Grade (percent of students)

Year	Mathematics	English	Science	History
2011	76	-	-	-
2009	77	-	-	-
2007	75	-	-	-
2005	73	-	-	-
2003	73	43	-	-
2000	73	-	26	-
1998	-	32	-	15
1996	71	35	21	-
1994	72	37	19	17
1992	73	50	-	-
1990	75	60	29	29

If the prevalence of grouping practices has changed and their nature evolved, it is likely that each of the aspects of grouping I have considered – influences on group assignment, group mobility rates, and the effects of grouping on learning behaviors – may have changed in important ways as well.

There is also a noticeable gap between third and eighth grades in the current project. I do not know how prevalent grouping practices, either within- or between-classes, might have been during these grades. And while I have made comparisons between my project and Kerckhoff's *Diverging Pathways*, it is important to remember that mine is not a proper panel study, but draws inferences from two different cohorts of students over two different periods of historical time.

While it is convenient to compare, for instance, the effect sizes of group assignment on learning behaviors in third and eighth grades for different sets of students, it is less useful than if we could compare the effects on the same group of students over time.

I have attempted to answer causal questions about the effects of group assignment by utilizing propensity-score-based stratification. The key assumptions of these methods – strong ignorability and the stable-unite treatment value assumption – are discussed in each chapter. For both set of analyses I have conducted (in primary and secondary schools), my findings are fairly robust to potential violations of the strong ignorability assumption. I was able to obtain high-quality balance between treated and control groups within strata on a large amount of strong predictor variables. Sensitivity analyses conducted after each propensity score analysis give some indication of the risks. Nonetheless, the possibility of unobserved confounding variables can never be completely discounted. The SUTVA is more problematic, as it is not hard to imagine how the treatment assignment of one student might affect the outcome of others, particular in the case of within-class grouping. It is unclear how violation of this assumption might bias my results.

Finally, the use of large-scale survey data has given me a great amount of statistical power to analyze small and subtle effects, as well as a nationally-representative sample. However, by limiting the sample to students that experienced grouped instruction and had data available on their group assignments, I have limited the generalizability of the sample to a national population. While I have good measures of student achievement and learning behaviors, limited sample sizes within classrooms and schools also means that I have an incomplete picture of how sampled students stood relative to their peers – a key aspect affecting both students' likely placements and how they will fare in different groups. The data also provides less fine-

grained information about how grouped instruction plays out in the classrooms. My literature review has suggested a number of ways in which group assignments can affect academic outcomes and learning behaviors (instructional differences, stigma or halo effects, peer effects, et cetera), but I cannot assess most of these with the available data and have not attempted to do so here.

Suggestions for Future Research

Fortunately, as least some of the limitations of this project can be overcome in future research. The National Center for Education Statistics is currently engaged in a replication of the ECLS-K 1998. The Early Childhood Longitudinal Study Kindergarten class of 2010 is now in its fifth year of data collection. Data on student's ability group placements, academic outcomes, and learning behaviors (using the same Approaches to Learning scale) have been collected in kindergarten, first, second, and third grades. It will be very interesting to compare how grouping practices and their effects have changed over the last fifteen years. The NCES has also recently completed another large-scale longitudinal study of high school students and their transitions to adulthood. The Educational Longitudinal Study of 2002 followed students from tenth grade through their experiences in postsecondary education and the labor market. It was my original intention to incorporate data from this study into this project, but as I delved into data several limitations became apparent. No data was collected before tenth grade, so a lack of critical information (particularly prior achievement and learning behaviors) made adequate propensity score analysis for that grade impossible. Also, while students were asked to report information on what high school program they were in during the tenth grade survey, this question was not included in the twelfth grade survey. This made direct comparison of mobility rates between

ELS and NELS impossible. Nonetheless, both datasets contain data on high school transcripts, so coursework-based definitions of group assignments can be constructed to allow us to understand how tracking has changed over the last few decades. As a matter of curiosity, I compared the student reports of track assignments for tenth grade between the two datasets. The column on the left is identical to the left side of Figure 8 in chapter 3, whereas the right side details comparable responses from the ELS dataset in 2002.

Table 25 – Student-reported Track Assignments in 1990 and 2002

Track	1990	2002
Academic/College Prep	40.5%	55.2%
General	48.8%	35.0%
Vocational	10.7%	9.8%

Tenth grade students in 2002 were more considerably more likely to report being in the academic track than in 1990, while the general and vocational tracks experienced declines in reported membership. But is this merely due to changes in student perceptions, or does it also reflect changes in coursework? How does a "college prep" high school program in 2002 compare to one in 1990? Does the decrease in exclusivity of the academic track change how it affects students learning behaviors? Drawing comparisons between older and newer datasets can provide important insights into how grouping practices have changed over time and how they might be best modified in the future.

Finally, while this current project was limited to using observational data, future work should (where possible) concentrate on identification strategies that exploit exogenous variation in group assignments. While observational studies have the advantage of considering outcomes that occur organically in traditional schooling situations, questions surrounding the effects of group assignment are plagued by the fact that treatments are allocated specifically on the basis of

expected outcomes. Research such as those following policy changes (Allensworth, Nomi, et al., 2009; Card and Giuliano, 2016), rule-based group assignments (Jackson, 2009), and opportunities for randomized controlled trials (Duflo et al., 2008) are necessary to validate findings from observational studies. Finally, and perhaps most importantly, we cannot confidently develop changes in grouping policies or practices until research (especially qualitative or mixed-methods research) that focuses on specific mechanisms provides us with a better understanding of how group assignments affects both learning behaviors and learning outcomes.

Appendix 1 – Descriptive Statistics for Selected ECLS-K Variables

Variable	Description	N	Mean	SD	Min	Max
	Outcomes					
ReadT1	Reading Achievement (1st grade)	13380	.15	.43	-1.97	1.44
LB1	Learning Behaviors (1st grade)	13380	3.06	.7	1	4
WkBest1	Works to Best of Ability (1st grade)	12930	3.17	.65	1	4
	Student-level covariates					
ReadTK	Reading Achievement (K)	13380	-.7	.49	-2.41	1.1
LBK	Learning Behaviors (K)	13380	3.15	.67	1	4
WkBestK	Works to Best of Ability (K)	13270	3.22	.63	1	4
Dmale	Male	13380	.51	.5	0	1
Dblack	Black	12940	.14	.35	0	1
Dhispan	Hispanic	12940	.13	.34	0	1
Dasian	Asian	12940	.05	.22	0	1
Dother	Other Race	12940	.04	.19	0	1
Dnoneng	Speaks other language at home	12930	.09	.29	0	1
Specneed	Has special needs	12570	.11	.31	0	1
Native	Native of US	12630	.98	.15	0	1
Momed1A	Mom has < HS diploma	12200	.1	.3	0	1
Momed1C	Mom has some college	10610	.27	.44	0	1
Momed1D	Mom is college grad	10610	.31	.46	0	1
Daded1A	Dad has < HS diploma	10050	.1	.3	0	1
Daded1C	Dad has some college	10050	.28	.45	0	1
Daded1D	Dad is college grad	10050	.32	.47	0	1
AgeEnt	Age entered K	11780	65.71	4.3	38	83
AgeFirst	Age at 1st non-parent care	10530	21.29	19.54	0	79
MomBAge	Mom's Birth Age	11570	24.06	5.47	12	46
MLang	Mother Language to child	11570	1.35	.85	1	4
DLang	Father language to child	9430	1.32	.82	1	4
Sibs1	Number of siblings	12250	1.49	1.11	0	11
Age1	Child age in months	13380	86.99	4.39	66.93	105.8
MscaleK	Math scale score	12190	2.2	7.34	6.9	59.82
GscaleK	Gen. know. scale score	11970	22.74	7.39	7.3	46.16
Height1	Height in inches	13070	48.45	2.39	35	59.5
Social1	Social skills index	12230	3.42	.53	1.33	4
BMI1	Body Mass Index	12720	16.87	2.85	5.61	44.35
NPCHrs1	Hrs in non-parent care	12190	5.44	8.97	0	140
ReadBoK	Freq. read books w/child	11770	3.28	.77	1	4
TellStK	Freq. tell stories	11770	2.75	.92	1	4
NumPlaK	# places lived	11720	2.06	1.27	1	20
ReadFrqK	Freq. reads outside school	12600	3	.99	1	4

MotorK	Motor skills	12070	12.38	3	0	17
MomAge	Mother's Age	11570	33.61	6.46	0	83
DadAge	Fathers Age	9430	36.56	6.8	1	85
IncomeK	Income	13010	55451	57995	0	999999
PraiseK	Freq. praise child	11770	1.2	.47	1	3
UpsetK	Freq upset with child	11770	2.75	.57	1	3
RdPicBk	Freq. read picture books	11750	3.32	.8	1	4
WicK	Family receives WIC	11690	.4	.49	0	1
YrsHome	Years in current home	7070	1.92	1.57	0	6
MuseumK	Freq. visit museum	12620	.31	.46	0	1
SportK	Freq. attends sports	12620	.45	.5	0	1
ZooK	Freq. goes to zoo	12620	.39	.49	0	1
NTrashK	Trash in area	12620	2.88	.37	1	3
NDrugK	Drugs in area	12600	2.88	.38	1	3
NSafe1	Neighborhood is safe	12230	2.73	.5	1	3
NBurgK	Burglaries in area	12590	2.87	.37	1	3
NViolK	Violence in area	12620	2.97	.21	1	3
NVacantK	Vacant homes in area	12620	2.94	.26	1	3
WarmK	Warm relationship with child	12420	1.3	.56	1	4
LikeK	Child likes R	12440	1.21	.47	1	4
TooBusy	Too busy for child	12440	3.44	.69	1	4
HrdWarm	Hard to be warm to child	12440	3.68	.65	1	4
BotherK	Child does things that bother	12430	3.3	.78	1	4
AffectK	Expresses physical affection	12440	1.11	.4	1	4
AngryK	Frq feel angry with child	12440	3.74	.48	1	4
SpankK	Frq child spanked	10050	.52	1.15	0	30
RCEver	Ever in relative's care	8990	.23	.42	0	1
Premat	Born premature	11640	.16	.37	0	1
FinProb	Financial problems in home	11720	.22	.42	0	1
KRepeat	Child repeated Kindergarten	13380	.04	.19	0	1
Insured	Has health insurance	12580	.93	.26	0	1
Headst	Was in Head Start	11770	.15	.35	0	1
Disab1	Disabled	12220	.16	.37	0	1
SingPar1	Single Parent	12250	.2	.4	0	1
MovedK1	Moved between K and 1st	13380	.07	.25	0	1
MTeachK	Parent met child's teacher	11790	.97	.16	0	1
ChoiceA1	Chose school	10410	.3	.46	0	1
ChoiceB1	Assigned school was choice	10410	.09	.28	0	1
Welfare1	Family on welfare	12200	.04	.2	0	1
Stamps1	Family on food stamps	12200	.12	.32	0	1
HomePC1	Computer in home	12240	.7	.46	0	1
Dance1	Participates in dance	12240	.21	.41	0	1
Athlet1	Participates in athletics	12240	.6	.49	0	1

Music1	Participates in music	12240	.12	.32	0	1
Crafts1	Participates in arts/crafts	12240	.11	.32	0	1
Perform1	Participates in performance arts	12240	.2	.4	0	1
AttOpenK	Attended open house	12610	.77	.42	0	1
AttPTAK	Attended PTA meeting	12620	.36	.48	0	1
AttPTCK	Attended PT conference	12620	.86	.35	0	1
AttEvtK	Attended school event	12620	.7	.46	0	1
VolK	Volunteered at school	12620	.52	.5	0	1
FundrK	Raised funds for school	12620	.64	.48	0	1
LibryK	Freq go to library	12620	.55	.5	0	1
SES1	Continuous Socioeconomic Status	12440	1.06	.79	7.04	12.88
AsTmAdj1	School days before assessment	13380	25.77	16.18	204	339
	Class-level covariates					
ClsPBlk1	Class percent black	12420	16.58	26.99	0	100
ClsPHis1	Class percent Hispanic	12420	11.27	21.22	0	100
Tage1	Teacher Age	12790	42.41	11.06	21	73
CSize1	Class Size	12550	2.84	4.64	1	52
CPBoys1	Class percent male	12230	.51	.28	0	9
CBehave1	Class behavior	12740	3.44	.9	1	5
CNew1	# of new kids in class	12240	.14	.24	0	1.83
CLeft1	Kids that left class	12290	.09	.12	0	1
CGifted1	Gifted Students	12150	.03	.08	0	1
CRdLo1	Class % reading below grade level	12190	.21	.17	0	2
CTardy1	Class tardiness	12340	.07	.07	0	1
CAbsen1	Class absences	12230	.04	.06	0	1
TDisTrn1	Trained to teach disabled	12310	3.16	1.05	1	5
TPPrep1	Paid prep hrs/wk	12250	1.95	.79	1	5
TUPrep1	Unpaid prep hrs/wk	12650	3.44	1.02	1	5
SchSprt1	School has spirit	12530	4.02	.85	1	5
Misbhv1	Misbehavior affects teaching	12510	2.25	1.08	1	5
NotCap1	Some kids can’t learn	12511	1.94	.94	1	5
Accept1	Staff accept T	12520	4.39	.67	1	5
PSupp1	Parent support	12490	3.82	.82	1	5
LowStds1	T thinks standards are low	12680	1.85	.84	1	5
Encour1	Admin encourages staff	12660	4.05	.99	1	5
TchAgn1	Would teach again	12670	4.32	.91	1	5
TWhite1	Teacher is white	12470	1.11	.31	1	2
TEd1	Teacher’s highest ed level	12600	4.11	.97	1	7
TCert1	Teachers’s certification type	12510	3.84	.78	1	5
CLDis1	Students with learning disability	7240	.04	.06	0	1
AssistA1	Assistance from teacher	12160	3.88	1.04	1	5
AssistB1	Assistance from tutors	11940	3.05	1.43	1	5

AssistC1	Assistance from specialist	11650	2.61	1.69	1	5
AssistD1	Assistance from pullout	11840	3.15	1.57	1	5
AssistE1	Assistance from other	12570	1.29	.92	1	5
PdAides1	# paid aids	12430	.67	.72	0	6
CMatsA1	Texbooks are adequate	11160	4.34	.77	2	5
CMatsB1	Tradebooks are adequate	12160	4.45	.71	2	5
CMatsC1	Workbooks are adequate	12500	4.08	.94	2	5
CMatsD1	Class space adequate	12710	4.08	.96	2	5
NumConf1	Number of PT conferences	12670	2.77	.68	1	4
TPrep1	Index of teacher prep	12450	2.9	1.16	0	5
PConf1	% parents attend conference	12690	4.46	.96	1	5
PHelp1	% parents volunteer	12700	2.31	.98	1	5
	School-level covariates					
SPHis1	School % Hispanic	10820	1.53	19.32	0	99.3
SPAsia1	School % Asian	10820	4.99	12.57	0	100
SPBlk1	School % black	10820	15.95	26.31	0	100
SAbsen1	Problems with absences	9330	1.92	.92	1	5
STurnov1	Problems with turnover	9320	1.81	.92	1	5
SActT1	Ts plan PD	9340	3.64	.97	1	5
SPDTime1	Adequate time for PD	9320	3.06	1.23	1	5
SIncent1	Incentives for improvement	9290	3.12	1.09	1	5
SUnion1	Union cooperates with admin	8970	3.64	1.01	1	5
SParInv1	Parents active	9310	4.03	.9	1	5
PrivSch1	Private School	13380	.16	.36	0	1
GrEnrl1	Grade-level enrollment	13020	79.97	48.19	0	374
SEnrl1	School enrollment	13290	3.36	1.14	1	5
SFrLun1	School % Free lunch	8840	25.96	27.74	0	100
STiI1	Title I program	11840	.66	.52	0	3
SPWht1	School % white	10820	65.02	34.52	0	100
SProDev1	Active PD program	9320	4.07	.89	1	5
STens1	Problems with racial tensions	11960	3.06	1.25	1	8
SCrime1	Problems with crime	11930	2.99	1.47	1	8
STheft1	Problems with theft	12080	1.93	.26	1	2
ScMatsA1	Library meets needs	11960	4.31	1.03	1	5
ScMatsB1	Classrooms meet needs	11950	4.44	.7	1	5
SPRead1	School Reading %ile	8490	65.06	22.63	0	100
SPMath1	School Math %ile	8420	65.48	22.63	0	100
SViol1	Problems with Violence	12030	1.68	.47	1	2
DisSrvA1	Disabled kids in reg class	13380	.26	.44	0	1
DisSrvB1	Disabled kids in separate cls	13380	.81	.39	0	1
SchMan1	School based management	12030	1.37	.48	1	2
SGdPr1	Personable Principal	12180	1.29	.62	1	4

SOrder1	Order in Hallways	12480	1.25	.56	1	4
SHelp1	Helpful staff	12570	1.21	.5	1	4
SSecur1	Adequate security	12530	1.17	.46	1	4
SHall1	Decorated halls	12300	1.28	.57	1	4
Sattent1	Attentive Teachers	12530	1.2	.45	1	4

Appendix 2 – Model 2 from Table 1.3 with Class Racial Composition

Level-1 Model

$\text{Prob}(HIGROUP1_{ijk}=1|\pi_{jk}) = \phi_{ijk}$
$\log[\phi_{ijk}/(1 - \phi_{ijk})] = \eta_{ijk}$
$\eta_{ijk} = \pi_{0jk} + \pi_{1jk}*(READTK_{ijk}) + \pi_{2jk}*(LBK_{ijk}) + \pi_{3jk}*(WKBESTK_{ijk}) + \pi_{4jk}*(HIGROUPK_{ijk})$
$+ \pi_{5jk}*(LOGROUPK_{ijk}) + \pi_{6jk}*(UNGROUPE_{ijk}) + \pi_{7jk}*(DMALE_{ijk}) + \pi_{8jk}*(DBLACK_{ijk})$
$+ \pi_{9jk}*(DHISPAN_{ijk}) + \pi_{10jk}*(DASIAN_{ijk}) + \pi_{11jk}*(DOTHER_{ijk}) + \pi_{12jk}*(DNONENG_{ijk})$
$+ \pi_{13jk}*(SPECNEED_{ijk}) + \pi_{14jk}*(SIBS1_{ijk}) + \pi_{15jk}*(DISAB1_{ijk}) + \pi_{16jk}*(SINGPAR1_{ijk}) + \pi_{17jk}*(ZSES1_{ijk})$

Level-2 Model

$\pi_{0jk} = \beta_{00k} + r_{0jk}$
$\pi_{1jk} = \beta_{10k}$
$\pi_{2jk} = \beta_{20k}$
$\pi_{3jk} = \beta_{30k}$
$\pi_{4jk} = \beta_{40k}$
$\pi_{5jk} = \beta_{50k}$
$\pi_{6jk} = \beta_{60k}$
$\pi_{7jk} = \beta_{70k}$
$\pi_{8jk} = \beta_{80k} + \beta_{81k}*(CLSPBLK1_{jk})$
$\pi_{9jk} = \beta_{90k}$
$\pi_{10jk} = \beta_{100k}$
$\pi_{11jk} = \beta_{110k}$
$\pi_{12jk} = \beta_{120k}$
$\pi_{13jk} = \beta_{130k}$
$\pi_{14jk} = \beta_{140k}$
$\pi_{15jk} = \beta_{150k}$
$\pi_{16jk} = \beta_{160k}$
$\pi_{17jk} = \beta_{170k}$

Level-3 Model

$\beta_{00k} = \gamma_{000} + u_{00k}$
$\beta_{10k} = \gamma_{100}$
$\beta_{20k} = \gamma_{200}$
$\beta_{30k} = \gamma_{300}$
$\beta_{40k} = \gamma_{400}$
$\beta_{50k} = \gamma_{500}$
$\beta_{60k} = \gamma_{600}$
$\beta_{70k} = \gamma_{700}$
$\beta_{80k} = \gamma_{800}$
$\beta_{81k} = \gamma_{810}$

$\beta_{90k} = \gamma_{900}$
$\beta_{100k} = \gamma_{1000}$
$\beta_{110k} = \gamma_{1100}$
$\beta_{120k} = \gamma_{1200}$
$\beta_{130k} = \gamma_{1300}$
$\beta_{140k} = \gamma_{1400}$
$\beta_{150k} = \gamma_{1500}$
$\beta_{160k} = \gamma_{1600}$
$\beta_{170k} = \gamma_{1700}$

READTK LBK WKBESTK HIGROUPK LOGROUPK UNGROUPE DMALE DBLACK DHISPAN DASIAN DOTHER DNONENG SPECNEED SIBS1 DISAB1 SINGPAR1 ZSES1 have been centered around the grand mean.

CLSPBLK1 has been centered around the grand mean.

Level-1 variance = $1/[\phi_{ijk}(1-\phi_{ijk})]$

Results for Non-linear Model with the Logit Link Function
Unit-Specific Model, PQL Estimation - (macro iteration 9)

τ_π
INTRCPT1,π_0 .44443
Standard error of τ_π
INTRCPT1,π_0 .08817

Random level-1 coefficient	Reliability estimate
INTRCPT1,π_0	.163

Note: The reliability estimates reported above are based on only 2600 of 5005 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

τ_β
INTRCPT1
INTRCPT2,β_{00}
.69329
Standard error of τ_β
INTRCPT1
INTRCPT2,β_{00}
.08810

Random level-2 coefficient	Reliability estimate

INTRCPT1/INTRCPT2, β_{00}	.341

Final estimation of fixed effects: (Unit-specific model)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, π_0					
For INTRCPT2, β_{00}					
INTRCPT3, γ_{000}	-1.169795	.052461	-22.299	1860	<.001
For READTK slope, π_1					
For INTRCPT2, β_{10}					
INTRCPT3, γ_{100}	3.154049	.098608	31.986	2005	<.001
For LBK slope, π_2					
For INTRCPT2, β_{20}					
INTRCPT3, γ_{200}	.395562	.068996	5.733	2005	<.001
For WKBESTK slope, π_3					
For INTRCPT2, β_{30}					
INTRCPT3, γ_{300}	.227414	.066887	3.400	2005	<.001
For HIGROUPK slope, π_4					
For INTRCPT2, β_{40}					
INTRCPT3, γ_{400}	1.024412	.102270	1.017	2005	<.001
For LOGROUPK slope, π_5					
For INTRCPT2, β_{50}					
INTRCPT3, γ_{500}	-.140766	.154291	-.912	2005	.362
For UNGROUPE slope, π_6					
For INTRCPT2, β_{60}					
INTRCPT3, γ_{600}	.499318	.093698	5.329	2005	<.001
For DMALE slope, π_7					
For INTRCPT2, β_{70}					
INTRCPT3, γ_{700}	-.057159	.059879	-.955	2005	.340
*For DBLACK slope, π_8					
For INTRCPT2, β_{80}					
INTRCPT3, γ_{800}	-.180280	.155536	-1.159	2005	.247
For CLSPBLK1, β_{81}					
INTRCPT3, γ_{810}	.014784	.003441	4.297	2005	<.001
For DHISPAN slope, π_9					
For INTRCPT2, β_{90}					
INTRCPT3, γ_{900}	.008299	.105551	.079	2005	.937
For DASIAN slope, π_{10}					
For INTRCPT2, β_{100}					
INTRCPT3, γ_{1000}	.143834	.148484	.969	2005	.333
For DOTHER slope, π_{11}					
For INTRCPT2, β_{110}					
INTRCPT3, γ_{1100}	.198758	.168494	1.180	2005	.238
For DNONENG slope, π_{12}					
For INTRCPT2, β_{120}					
INTRCPT3, γ_{1200}	.345761	.119572	2.892	2005	.004

For SPECNEED slope, π_{13}					
For INTRCPT2, β_{130}					
INTRCPT3, γ_{1300}	-.039985	.112828	-.354	2005	.723
For SIBS1 slope, π_{14}					
For INTRCPT2, β_{140}					
INTRCPT3, γ_{1400}	-.024942	.027338	-.912	2005	.362
For DISAB1 slope, π_{15}					
For INTRCPT2, β_{150}					
INTRCPT3, γ_{1500}	.012110	.090427	.134	2005	.893
For SINGPAR1 slope, π_{16}					
For INTRCPT2, β_{160}					
INTRCPT3, γ_{1600}	-.074240	.083405	-.890	2005	.374
For ZSES1 slope, π_{17}					
For INTRCPT2, β_{170}					
INTRCPT3, γ_{1700}	-.074594	.036889	-2.022	2005	.043

Appendix 3 – Descriptive Statistics for Selected NELS Variables

Variable	Description	N	Mean	Std Dev	Min	Max
Math8	Math IRT theta – 8th grade	10970	45.48	8.67	23.98	67.23
Math1	Math IRT theta – 10h grade	10810	5.73	9.88	24.87	72.9
Math2	Math IRT theta – 12th grade	9150	54.30	1.48	26.3	8.67
Read8	Reading IRT theta – 8th grade	10960	46.74	8.61	23.96	63.49
Read1	Reading IRT theta – 10h grade	10820	5.32	1.18	22.18	71.86
Read2	Reading IRT theta – 12th grade	9140	53.03	1.70	23.2	78.51
LB8	Learning behaviors – 8th grade	10970	3.57	.48	1	5
LB1	Learning behaviors – 10th grade	11130	3.26	.47	1	4
LB2	Learning behaviors – 12th grade	10660	3.41	.60	1	5
LBT8	Teacher-report of LB – 8th grade	10250	.86	.24	0	1
LBT1	Teacher-report of LB – 10th grade	9660	4.01	.58	1	5
LBT2	Teacher-report of LB – 12th grade	6220	3.98	.63	1	5
AcaTr1	Academic Track – 10th grade	10800	.40	.48	0	1
AcaTr2	Academic Track – 12th grade	11920	.45	.49	0	1
GenTr1	General Track – 10th grade	10790	.48	.50	0	1
GenTr2	General Track – 12th grade	11920	.42	.49	0	1
VocTr1	Vocational Track – 10th grade	10790	.10	.30	0	1
VocTr2	Vocational Track – 12th grade	11920	.12	.32	0	1
UpMob81	Upwardly Mobile – 8th to 10th grade	10120	.20	.40	0	1
DnMob81	Downwardly Mobile – 8th to 10th grade	10120	.14	.35	0	1
UpMob12	Upwardly Mobile – 10th to 12th grade	10630	.14	.35	0	1
DnMob12	Downwardly Mobile – 10th to 12th grade	10630	.08	.27	0	1
Dropout1	Dropped out between 8th and 10th	12050	.04	.19	0	1
Dropout2	Dropped out between 10th and 12th	12040	.08	.27	0	1
EnCollege	Enrolled in bachelor degree program	12000	.36	.48	0	1
EnPSE	Enrolled in other postsec ed program	12000	.68	.46	0	1
FTJob	Had Full-time job	12140	.76	.42	0	1
WeeksWk	Weeks worked in prior year	10810	46.72	11.06	0	52

Income	Income in prior year	11150	24942.35	2019.49	0	500000
WageRate	Hourly Wage Rate	11400	14.47	11.04	.001	25.00
HiGroup8	High Ability Group in 8th grade	11380	.33	.47	0	1
MdGroup8	Middle Ability Group in 8th grade	11380	.41	.49	0	1
LoGroup8	Low Ability Group in 8th grade	11380	.07	.25	0	1
Ungroup8	Not Ability Grouped in 8th grade	11380	.20	.40	0	1
Male	S is male	11300	.47	.50	0	1
Dwhite	S is white	11300	.71	.45	0	1
Dblack	S is black	11300	.09	.29	0	1
Dhispan	S is Hispanic	11300	.13	.33	0	1
Dasian	S is Asian	11300	.07	.25	0	1
Sibs	# of Siblings	11320	2.27	1.57	0	6
HomePC	Family owns PC	10950	.43	.49	0	1
OwnRoom	S has own room	12140	.75	.43	0	1
Chores	Freq P require chores done	11310	1.47	.73	1	4
TGood	S reports teaching is good	11120	2.03	.70	1	4
Gifted	S classified as gifted	10920	.20	.40	0	1
TimeAlon	Time S spends alone after school	11200	1.76	1.21	0	4
Grades8	Grade 8 grades composite	11290	2.98	.73	.5	4
Sports	S participates in sports	10520	.44	.50	0	1
SelfCon8	Academic self-concepts – 8th grade	11320	.01	.65	-2.91	1.23
HomeLang	Non-English Home	11360	3.62	.80	1	4
PHSGrad	Parent is High school grad	10710	.64	.48	0	1
PColGrad	Parent is college grad	10710	.25	.43	0	1
Alone	S alone when gets home from school	11200	1.76	1.21	0	4
Retaind8	S Retained in 8th grade	11470	.14	.34	0	1
Discuss	Frequency P discusses school with S	11230	2.43	.67	1	3
SpokeT	Frequency P speaks with teacher	9880	.65	.48	0	1
CheckHW	Frequency P checks homework	11320	1.93	1.00	1	4
Trust	Parents trust S to do what they say	11280	.80	.40	0	1
Office	S sent to office for misbehaving	11240	.35	.62	0	2

SES1	Socioeconomic composite – 10th grade	11250	-.04	.81	-2.838	2.762
Moved1	Moved in 2 years before 10th grade	10760	.15	.36	0	1
Married1	Parent married in 2 years before 10th grade	10760	.05	.22	0	1
Divorced1	Parent divorced 2 years before 10th grade	10760	.07	.25	0	1
BSit1	Time S spends sitting siblings	10140	.38	.49	0	1
SUrban8	School Urbanicity – 8th grade	11380	1.94	.75	1	3
SPLunch8	School % FR lunch – 8th grade	11180	22.47	22.75	0	100
SPMin8	School % minority – 8th grade	5200	2.60	2.49	0	8
SPublic1	Schools is public – 10th grade	10660	.87	.34	0	1
SSinPar1	School % single parents – 10th grade	9140	1.64	.65	0	5
SPLEP1	School % LEP – 10th grade	10530	1.06	1.33	0	9
SPLunch1	School % free-reduced lunch – 10th grade	9950	19.32	21.02	0	100
SPWhite1	School % white – 10th grade	10370	73.08	3.36	0	100
SPRead1	School % in remedial reading classes	10160	8.00	9.40	0	100
SPMath1	School % in remedial math classes	10010	8.13	9.73	0	100
SPTA1	School has PTA	9000	1.39	.49	1	2
STMoral1	School teacher morale	9120	3.86	.79	1	5
SSMoral1	School student morale	9150	3.96	.67	1	5
SAbsent1	Absenteeism problem for school	9160	2.48	.83	1	4
SFight1	Fights problem for school	9160	1.80	.64	1	4
STPrep1	Teacher prep in school	9110	57.85	2.19	0	200
SDOProg1	Drop out program n school	10510	.57	.50	0	1
SPVol1	% parents volunteering for school	8980	12.46	15.35	0	100
SMagnet1	School is magnet	10590	.05	.23	0	1
SCath1	School is Catholic	10590	.04	.19	0	1
SChoice1	Public school of choice	10590	.15	.36	0	1
SUrban1	School urbanicity	11770	1.97	.77	0	3
SPublic2	School is public	9710	.79	.41	0	1
SSinPar2	School % single parents	9040	2.53	.83	1	5
SPLEP2	School % LEP – 12th grade	9660	.83	1.19	0	9
SPLunch2	School % free-reduced lunch – 12th grade	9560	2.47	2.90	0	100

SPWhite2	School % white – 12th grade	9620	73.61	29.88	0	100
SPRead2	School % in remedial reading classes	9600	7.27	9.91	0	100
SPMath2	School % in remedial math classes	9580	7.69	1.02	0	75
STMoral2	School teacher morale – 12th grade	8960	2.32	.56	1	3
SSMoral2	School student morale – 12th grade	8960	2.40	.54	1	3
SMagnet2	School is magnet – 12th grade	9580	.09	.28	0	1
SChoice2	Public school of choice – 12th grade	9570	.27	.44	0	1
SCath2	School is Catholic – 12th grade	9560	.07	.25	0	1
SPPTA2	School use PTA – 12th grade	8340	1.91	1.11	1	5
SDOProg2	School has dropout program	8720	1.45	.50	1	2
SPVol2	Proportion of parents volunteering	8610	1.35	.75	1	5
SAbsent2	Absenteeism problem in school	8860	2.58	.86	1	4
SFight2	Fights problem in school	8880	3.23	.68	1	4
SUrban2	School urbanicity	11680	1.97	.77	1	3

Appendix 4 – Reliability and Inter-correlation Information for Learning Behavior Outcomes

Student Reports

My indices of learning behaviors are all simple means of survey items. The scale of responses (Yes/No, 1-4 ordinal, 1-5 ordinal) varies from index to index, but with each index, all items generally have the same scale (the 8th and 12th grade student reports of learning behaviors do have some items with both 4- and 5- point ordinal scales) . Items were reverse-coded as necessary so that higher values represented greater frequency of positive behaviors or lower frequency of negative behaviors (e.g. cutting class). Within each index, all correlations are significant with $p<.001$. If data was missing for individual items, index values were calculated from the remaining items. Imputation was performed on the index values, not each item individually.

The eighth-grade student-reported measure of learning behaviors is the mean of five items from the 1988 student survey.

Cronbach's alpha = .632.

Survey Item	**Variable**	BYS76	BYS77	BYS78A	BYS78B	BYS78C
Skip/cut classes	BYS76	1	.232	.162	.180	.162
Late for school	BYS77		1	.173	.178	.178
Come w/o pencil/paper	BYS78A			1	.468	.398
Come w/o books	BYS78B				1	.443
Come w/o homework	BYS78C					1

The 10th grade student-reported measure of learning behaviors is the mean of eight items from the 1990 student survey. Technically items F1S12A through E ask students about their mindsets, how they *feel* about certain behaviors. However, they refer to feelings about specific outward behaviors, so I decided to include them in my analysis. The survey included similar items (e.g. "It's okay to work hard for good grades.") but responses were limited to Yes/No, and did not discriminate well between students. I also did not include questions that asked about general behaviors not related to learning (e.g. how many times the respondent got in trouble or was suspended).

Cronbach's alpha = .769

Survey Item	**Variable**	F1S12A	F1S12B	F1S12C	F1S12D	F1S12E	F1S40A	F1S40B	F1S40A
Late for class	F1S12A	1	.476	.414	.276	.313	.183	.192	.215
Cut a couple classes	F1S12B		1	.587	.387	.359	.181	.227	.234
Skip school for a day	F1S12C			1	.342	.338	.163	.186	.210
Cheat on tests	F1S12D				1	.618	.171	.146	.177
Copy homework	F1S12E					1	.185	.156	.211
Come w/o pen/paper	F1S40A						1	.542	.356
Come w/o books	F1S40B							1	.392
Come w/o homework	F1S40A								1

The 12th-grade measure of student-reported learning behaviors is the mean of seven items from the 1992 student survey. Questions F2S21A through D asked students specifically about their behaviors in their current or most recent math class, which questions F2S24A through C asked students about their behaviors generally. Cronbach's alpha = .715

Survey Item	**Variable**	F2S24A	F2S24B	F2S24C	F2S21A	F2S21B	F2S21C	F2S21D
Come w/o pen/paper	F2S24A	1	.600	.399	.203	.191	.086	.070
Come w/o books	F2S24B		1	.413	.197	.181	.084	.103
Come w/o homework	F2S24C			1	.232	.343	.179	.164
Pay attention	F2S21A				1	.507	.313	.455
Complete work on time	F2S21B					1	.349	.369
Does more work than req'd	F2S21C						1	.381
Participates in class	F2S21D							1

Teacher Reports

The 8th grade teacher report of learning behaviors of sampled students is the mean of five items from the 1988 teacher survey.

Cronbach's alpha = .749

Survey Item	**Variable**	BYT4_2	BYT4_3	BYT4_4	BYT4_5	BYT4_6
Performs below ability	BYT4_2	1	.645	.269	.231	.554
Rarely completes homework	BYT4_3		1	.29	.288	.555
Frequently absent	BYT4_4			1	.286	.216
Frequently tardy	BYT4_5				1	.291
Inattentive in class	BYT4_6					1

The 10th grade teacher report of learning behaviors of sampled students is the mean of eight items from the 1990 teacher survey. In this year, two teachers were sampled for each student. If data was only available for one teacher, the index was formed using only that data. Each teacher answered four questions concerning sampled students' behaviors in their class.

Cronbach's alpha = .822.

Survey Item	**Variable**	F1T1_15	F1T1_16	F1T1_17	F1T1_18	F1T5_15	F1T5_16	F1T5_17	F1T5_18
Does Homework	F1T1_15	1	.429	.378	.671	.473	.295	.261	.399
Absent	F1T1_16		1	.379	.387	.298	.416	.217	.238
Tardy	F1T1_17			1	.397	.264	.229	.331	.238
Attentive in class	F1T1_18				1	.398	.260	.253	.377
Does Homework – T2	F1T5_15					1	.418	.392	.670
Absent – T2	F1T5_16						1	.351	.353
Tardy – T2	F1T5_17							1	.388
Attentive in class – T2	F1T5_18								1

The 12th grade teacher report of learning behaviors of sampled students is the mean of four items from the 1992 teacher survey. These are the same as the 1990 survey, but only one teacher was sampled in 1992.

Cronbach's alpha = .762.

Survey Item	**Variable**	F2T1_12	F2T1_13	F2T1_14	F2T1_15
Does Homework	F2T1_12	1	.437	.383	.653
Absent	F2T1_13		1	.421	.378
Tardy	F2T1_14			1	.406
Attentive in class	F2T1_15				1

Appendix 5 –Table 14.1, Model 3. Key Student- and School-Level Covariates Predicting Academic Track Membership in 10th Grade

Level-1 Model

$\text{Prob}(ACATR1_{ij}=1|\beta_j) = \phi_{ij}$

$\log[\phi_{ij}/(1 - \phi_{ij})] = \eta_{ij}$

$\eta_{ij} = \beta_{0j} + \beta_{1j}*(HIGROUP8_{ij}) + \beta_{2j}*(LOGROUP8_{ij}) + \beta_{3j}*(UNGROUP8_{ij}) + \beta_{4j}*(LB8_{ij}) + \beta_{5j}*(LBT8_{ij}) + \beta_{6j}*(READ8_{ij}) + \beta_{7j}*(MATH8_{ij}) + \beta_{8j}*(MALE_{ij}) + \beta_{9j}*(DBLACK_{ij}) + \beta_{10j}*(DHISPAN_{ij}) + \beta_{11j}*(DASIAN_{ij}) + \beta_{12j}*(SIBS_{ij}) + \beta_{13j}*(GIFTED_{ij}) + \beta_{14j}*(HOMELANG_{ij}) + \beta_{15j}*(SES1_{ij}) + \beta_{16j}*(RETAIND8_{ij}) + \beta_{17j}*(DISCUSS_{ij}) + \beta_{18j}*(MOVED1_{ij}) + \beta_{19j}*(MARRIED1_{ij}) + \beta_{20j}*(DIVORCED_{ij})$

Level-2 Model

$\beta_{0j} = \gamma_{00} + \gamma_{01}*(SSINPAR1_j) + \gamma_{02}*(SPLEP1_j) + \gamma_{03}*(SPLUNCH1_j) + \gamma_{04}*(SPWHITE1_j) + \gamma_{05}*(SPVOL1_j) + \gamma_{06}*(SMAGNET1_j) + \gamma_{07}*(SCATH1_j) + \gamma_{08}*(SCHOICE1_j) + u_{0j}$

$\beta_{1j} = \gamma_{10}$

$\beta_{2j} = \gamma_{20}$

$\beta_{3j} = \gamma_{30}$

$\beta_{4j} = \gamma_{40}$

$\beta_{5j} = \gamma_{50}$

$\beta_{6j} = \gamma_{60}$

$\beta_{7j} = \gamma_{70}$

$\beta_{8j} = \gamma_{80}$

$\beta_{9j} = \gamma_{90}$

$\beta_{10j} = \gamma_{100}$

$\beta_{11j} = \gamma_{110}$

$\beta_{12j} = \gamma_{120}$

$\beta_{13j} = \gamma_{130}$

$\beta_{14j} = \gamma_{140}$

$\beta_{15j} = \gamma_{150}$

$\beta_{16j} = \gamma_{160}$

$\beta_{17j} = \gamma_{170}$

$\beta_{18j} = \gamma_{180}$

$\beta_{19j} = \gamma_{190}$

$\beta_{20j} = \gamma_{200}$

Works Cited

Ainsworth-Darnell, James, and Douglas Downey. 1998. “Assessing the Oppositional Culture Explanation for Racial/Ethnic Differences in School Performance.” *American Sociological Review* 62: 536-53.

Alexander, Karl, Martha Cook, and Edward McDill. 1978. "Curriculum Tracking and Educational Stratification: Some Further Evidence." *American Sociological Review* 43:47-66.

Allensworth, Elaine, Takao Nomi, Nicholas Montgomery, and Valerie Lee. 2009. "College Preparatory Curriculum for All: Academic Consequences of Requiring Algebra and English I for Ninth Graders in Chicago." *Educational Evaluation and Policy Analysis* 31(4): 367-391.

Allington, Richard. 1983. “The reading instruction provided readers of differing reading ability.” *Elementary School Journal* 83(5): 548-559.

Ansalone, George. 2001. "Schooling, tracking, and inequality." *Journal of Children & Poverty* 7(1): 33-47.

Ansalone, George. 2003. "Poverty, Tracking, and the social construction of failure: international perspectives on tracking." *Journal of Children & Poverty* 9(1): 3-20.

Argys, Laura, Daniel Rees, and Dominic Brewer. 1996. "Detracking America’s schools: equity at zero cost?" *Journal of Policy Analysis and Management* 15(4): 623-45.

Austin, Mary, and Coleman Morrison. 1963. *The first R: The Harvard Report on reading in elementary schools.* New York: Macmillan.

Barr, Rebecca, and Robert Dreeben. 1983. *How Schools Work*. Chicago: University of Chicago Press.

Berends, Mark. 1995. "Educational Stratification and Students' Social Bonding to School." *British Journal of Sociology of Education* 16(3): 327-51.

Betts, Julian and Jaime Shkolnik. 2000. “Key Difficulties in Identifying the Effects of Ability Grouping on Student Achievement.” *Economics of Education Review,* February, 19(1): 21-26.

Billett, Roy. 1932. *The Administration and Supervision of Homogeneous Grouping.* Columbus: Ohio State University Press.

Blauner, Robert. 1964. *Alienation and Freedom*. Chicago: University of Chicago Press.

Bowles, Simon and Herbert Gintis. 1976. *Schooling in Capitalist America*. New York: Basic Books.

Burris, Carol. 2014. *On the Same Track: How Schools Can Join the Twenty-First-Century Struggle Against Resegregation*. Boston: Beacon Press.

Carbonaro, William, and Adam Gamoran. 2002. "The Production of Achievement Inequality in High School English." *American Educational Research Journal* 39(4): 801-27.

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

Card, David and Laura Giuliano. 2016. "Can Tracking Raise the Test Scores of High-Ability Minority Students?" *NBER Working Paper* 22104. National Bureau of Economic Research.

Carey, Nancy, Elizabeth Farris, and Judi Carpenter. 1994. *Curricular Differentiation in Public High Schools*. Washington DC: National Center for Education Statistics.

Catsambis, Sophia, and Anthony Buttaro Jr. 2012. "Revisiting 'Kindergarten as Academic Boot Camp': A Nationwide Study of Ability Grouping and Pyscho-Social Development." *Social Psychology of Education* 15(4) 483-515.

Chorzempa, Barbara, and Steve Graham. 2006. "Primary-grade teachers' use of within-class ability grouping in reading." *Journal of Educational Psychology* 98(3): 529–541.

Chui, Donna, Yodit Beru, Erin Watley, Salen Wubu, Emma Simson, Robin Kessenger, Anahi Rivera, Patrick Schmidlein, Allan Wigfield. 2008. "Influence of Math Tracking on Seventh Grade Students' Self-Beliefs and Social Comparisons." *Journal of Educational Research* 102(2): 125-135.

Clark-Ibanez, Marisol. 2005. "Making Meaning of Ability Grouping in Two Urban Schools." *International Review of Modern Sociology* 31(1): 57-79.

Condron, Dennis. 2007. "Stratification and educational sorting: Explaining ascriptive inequalities in early childhood reading group placement." *Social Problems,* 54(1):139-160.

Condron, Dennis. 2008. "Skill grouping and unequal reading gains in the elementary years." *Sociological Quarterly*, 49(2): 363–394.

Cooper, Robert. 1996. "Detracking Reform in an Urban California High School: Improving the Schooling Experiences of African American Students." *Journal of Negro Education* 62(2): 190-208.

Cremin, Lawrence. 1961. *The Transformation of the School*: *Progressivism in American Education.* New York: Knopf.

DeLany, Brian. 1991. "Allocation, Choice, and Stratification within High Schools: How the Sorting Machine Copes." *American Journal of Education* 9(2):181-207.

Dreeban, Robert. 1984. "First-Grade Reading Groups: Their Formation and Change." In *The Social Context of Instruction* , edited by Penelope Peterson, Louise Wilkinson, and Maureen Hallinan, 69-84. Orlando, FL: Academic Press.

Dreeben, Robert and Rebecca Barr. 1988. "Classroom Composition and the Design of Instruction." *Sociology of Education* 61(3): 129-142.

Dreeben, Robert. 1968. *On What is Learned in Schools*. Reading, Massachusettes: Addison-Wesley Publishing Company.

Drews, Elizabeth. 1963. *Student abilities, grouping patterns, and classroom interaction.* Cooperative Research Project No. 608. Washington DC: U.S. Department of Health, Education, and Welfare.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2008. "Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *NBER Working Paper* 14475. National Bureau of Economic Research.

Dweck, Carol. 2001. "The Development of Ability Conceptions." In *Development of Achievement Motivation,* edited by Allan Wigfield and Jacquelynne Eccles, 57-83. San Diego, CA: Academic Press.

Eder, Donna and Diane Felmee. 1984. "The Development of Attention Norms in Ability Groups." In *The Social Context of Instruction,* edited by Penelope Peterson, Louise Wilkinson, and Maureen Hallinan, 69-84. Orlando, FL: Academic Press.

Eder, Donna. 1981. "Ability Grouping as a Self- Fulfilling Prophecy: A Micro Analysis of Teacher-Student Interaction." *Sociology of Education* 54(3):151-62.

Eder, Donna. 1983. "Organizational Constraints and Individual Mobility: Ability Group Formation and Maintenence." *Sociological Quarterly* 24(3): 405-420.

Farrington, Camille, Melissa Roderick, Elaine Allensworth, Jenny Nagaoka, Tasha Seneca Keyes, David Johnson, and Nicole Beechum. 2012. *Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance: A Critical Literature Review.* Chicago: Consortium for Chicago School Research.

Felmlee, Diane, and Donna Eder. 1983. "Contextual effects in the classroom: The impact of ability groups on student attention." *Sociology of Education* 56(2): 77‑87.

Ferguson, Ronald. 2007. *Toward Excellence with Equity*: *An Emerging Vision for Closing the Achievement Gap.* Cambridge, MA: Harvard Education Press.

Figlio, David, and Marianne Page. 2002 "School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality?" *Journal of Urban Economics* 51(3): 497-514.

Finley, Donna. 1984. "Teachers and Tracking in a Comprehensive High School." *Sociology of Education* 57(4): 233-243.

Gamoran Adam, and Matthew Weinstein. 1998. "Differentiation and opportunity in restructured schools." *American Journal of Education* 106(3): 385-431.

Gamoran, Adam, and Robert Mare. 1989. "Secondary School Tracking and Educational Inequality: Compensation, Reinforcement, or Neutrality?" *American Journal of Sociology* 94(5):1146-83.

Gamoran, Adam. 1987. "Organization, instruction, and the effects of ability grouping: Comment on Slavin's 'best-evidence synthesis.'" *Review of Educational Research,* 57(3): 341–345.

Gamoran, Adam. 1989. "Rank, Performance, and Mobility in Elementary School Grouping." *The Sociological Quarterly* 30(1): 109-123.

Gamoran, Adam. 1992. "The Variable Effects of High School Tracking." *American Sociological Review* 57: 812-828.

Garet, Michael and Brian DeLany. 1988. "Student, Courses, and Stratification." *Sociology of Education* 61(2):61-77.

Gouldner, Helen. 1978. *Teacher's Pets, Troublemakers, and Nobodies*. Westport, CT: Greenwood Press.

Grant, Linda, and James Rothenberg. 1986. "The social enhancement of ability differences: Teacher-student interactions in first- and second-grade reading groups." *Elementary School Journal* 87(1): 29–50.

Grigg, Jeffrey. 2012. "School enrollment changes and student achievement growth: A case study in education disruption and continuity." *Sociology of Education* 85(4): 388-404.

Groff, Partrick. 1962. "A survey of basal reading grouping practices." *The Reading Teacher*: 15(4): 232-235.

Haller, Emil, and Margaret Waterman. 1985. "The criteria of reading group assignments." *Reading Teacher* 38(8): 772-782.

Hallinan, Maureen, and Aage Sørenson. 1983. "The Formation and Stability of instruction Groups." *American Sociological Review* 48(6): 838-851.

Hallinan, Maureen, and Aage Sørenson. 1986. "Student Characteristics and Assignment to Ability Groups: Two Conceptual Formulations." *Sociological Quarterly* 27(1):1-13.

Hallinan, Maureen. 1994. “Tracking: From Theory to Practice.” *Sociology of Education* 67(2): 79-91.

Hallinan, Maureen. 1996. “Tracking Mobility in Secondary School.” *Social Forces* 74(3): 983-1002.

Hanushek, Eric and Ludger Wößmann. 2006. “Does Educational Tracking Affect Performance and Inequality?” *The Economic Journal* 116(510): C63-C76.

Hanushek, Eric, John Kain, and Steven Rivkin. 2004. “Disruption versus Tiebout improvement: The costs and benefits of switching schools.” *Journal of Public Economics* 88: 1721-1746.

Hawkins, Michael. 1966. “Mobility of students in reading groups.” *The Reading Teacher* 20(2): 136-140.

Heck, Ronald, Carol Price, and Scott Thomas. 2004. "Tracks as emergent structures: A network analysis of student differentiation in a high school." *American Journal of Education* 110(4): 321–353.

Hoffer, Thomas. 1992. "Middle School Ability Grouping and Student Achievement in Science and Mathematics." *Educational Evaluation and Policy Analysis* 14(3): 205-27.

Hong, Guangeli, Carl Corter, Yihua Hong, and Janette Pelletier. 2012. "Differential effects of literacy instruction time and homogeneous grouping in kindergarten: Who will benefit? Who will suffer?" *Educational Evaluation and Policy Analysis* 34(1): 69-88.

Hong, Guanglei, and Yihua Hong. 2009. "Reading instruction time and homogeneous grouping in kindergarten: An application of marginal mean weighting through stratification." *Educational Evaluation and Policy Analysis* 31(1): 54-81.

Huang, Min-Hsiung. 2009. "Classroom homogeneity and the distribution of student math performance: A country-level fixed-effects analysis" *Social Science Research* 38: 781-91.

Ireson, Judith and Susan Hallam. 2001. *Ability Grouping in Education*. Thousand Oaks, CA: Sage Publications Inc.

Jackson, C.Kirabo. 2009. "Ability-grouping and Academic Inequality: Evidence from Rule-Based Student Assignment." *NBER Working Paper* 14911. National Bureau of Economic Research.

Jones, James, Beth Vanfossen, and Margaret Ensminger. 1995. "Individual and Organizational Predictors of High School Track Placement." *Sociology of Education* 68(4): 287-300.

Kelly, Sean and Heather Price. 2011. “The Correlates of Tracking Policy: Opportunity Hoarding, Status Competition, or a Technical-Functional Explanation?” *American Educational Research Journal* 48(3): 560-585.

Kerckhoff, Alan. 1993. *Diverging Pathways: Social structure and career deflections*. Cambridge: Cambridge University Press.

Kulik Chen-Lin, and James Kulik. 1982. "Research synthesis on ability grouping." *Educational Leadership* 39(8): 619–21.

Kulik, James. 1992. *An analysis of the research on ability grouping: Historical and contemporary perspectives* (RBDM 9204). Storrs, CT: The National Research Center on the Gifted and Talented, University of Connecticut.

Lee, Valerie, and Anthony Bryk. 1988. "Curriculum Tracking as Mediating the Social Distribution of High School Achievement." *Sociology of Education* 61(2): 78-94.

Lewis, Amanda and John Diamond. 2015. *Despite the Best Intentions: How Racial Inequality Thrives in Good Schools.* Oxford, UK: Oxford University Press.

Lleras, Christy, and Claudia Rangel. 2009. "Ability grouping practices in elementary school and African American/Hispanic achievement." *American Journal of Education* 115(2): 279–304.

Lou, Yiping, Philip Abrami, and John Spence. 2000. "Effects of within-class ability grouping on student achievement: An exploratory model*." The Journal of Educational Research* 94(2): 101–112.

Lou, Yiping, Philip Abrami, John Spence, Catharine Poulsen, Bette Chambers, and Syvia d’Apollonia. 1996. "Within class grouping: A meta-analysis." *Review of Educational Research* 66(4): 423–458.

Loveless, T. 1999. *The Tracking Wars: State Reform Meets School Policy*. Washington, DC: Brookings Institution Press.

Loveless, Tom. 2013. *How Well Are American Students Learning*. Washington, DC: Brown Center on Education Policy, the Brookings Institution.

Lucas, Samuel and Aaron Good. 2001. "Race, class, and tournament track mobility." *Sociology of Education* 74(2): 139-156.

Lucas, Samuel and Mark Berends. 2002. "Sociodemographic Diversity, Correlated Achievement, and De Facto Tracking." *Sociology of* Education 75(4): 328-348.

Lucas, Samuel. 1999. *Tracking Inequality: Stratification and Mobility in American High Schools.* New York: Teachers College Press.

Lucas, Samuel. 2001. "Effectively maintained inequality: educational transitions, track mobility, and social background effects." *American Journal of Sociology* 106(6): 1642-1690.

Marascuilo, Leonard and Maryellen McSweeney. 1972. "Tracking and minority student attitudes and performance." *Urban Education* 6(4): 303-19.

Marsh, Herbert. 1987. "The big-fish-little-pond effect of academic self-concept." *Journal of Educational Psychology*, 79(3): 280–295.

McFarland, Daniel and Simon Rodan. 2009. "Organization by Design: Supply- and Demand-side Models of Mathematics Course Taking." *Sociology of Education* 82(4): 315- 343.

Meyer, John and Brian Rowan. 1977. "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *American Journal of Sociology* 83(2): 340-363.

Mickelson, Roslyn, and Bobbie Everett. 2008. "Neotracking in North Carolina: How high school courses of study reproduce race and class-based stratification." *Teachers College Record* 110(3): 535–570.

Moller, Stephanie and Elizabeth Stearns. 2012. "Tracking Success: High School Curricula and Labor Market Outcomes by Race and Gender." *Urban Education* 47(6): 1025-1-54.

Mulkey, Lynn, Sophia Catsambis, Lala Steelman, and Robert Crain. 2005. "The long-term effects of ability grouping in mathematics: a national investigation." *Social Psychology of Education* 8:137–77.

National Education Association Website. 2016 "Research Spotlight on Academic Ability Grouping." National Education Association. Accessed July 10, 2016. http://www.nea.org/tools/16899.htm

Nomi, Takako. 2010. "The Effects of Within-Class Ability Grouping on Academic Achievement in the Early Elementary Years." *Journal of Research on Educational Effectiveness* 3(1): 56-92.

Oakes, Jeannie. 2005. *Keeping Track: How Schools Structure Inequality*: *2nd Edition.* New Haven, CT: Yale University Press.

Oakes, Jeannie. and Gretchen Guiton. 1995. "Matchmaking: The Dynamics of High School Tracking Decisions." *American Educational Research Journal* 32(1): 3-33.

Page, Reba. 1991. *Lower Track Classrooms: A Curricular and Cultural Perspective*. New York: Teachers College Press.

Pallas, Aaron, Doris Entwisle, Karl Alexander, and Maria Stluka. 1994. "Ability-Group Effects: Instructional, Social, or Institutional?" *Sociology of Education* 67(1): 27-46.

Parsons, Talcott. 1959. “The School Class as a Social System: Some of Its Functions in American Society. *Harvard Education Review* 29(4): 297-318.

Pink, William and Mary Ellen Sweeney. 1978. “Teacher Nomination, Deviant Career Lines, and the Management of Stigma in the Junior High School.” *Urban Education* 13(3): 361-380.

Raudenbush, Stephen and Sally Sadoff. 2008. “Statistical Inference when classroom quality is measured with error.” *Journal of Research on Educational Effectiveness* 1(2): 138-154.

Raudenbush, Stephen, Brian Rowan, and Yuk Fai Cheong. 1993. “Higher Order Instructional Goals in Secondary Schools: Class, Teacher, and School Influences” *American Educational Research Journal* 20(3): 523-53.

Reardon, Sean. 2011. “The widening academic achievement gap between the rich and the poor: New evidence and possible explanations.” In *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children*, edited by Richard Murnane and Greg Duncan, 91-116. New York: Russell Sage Foundation Press.

Reuman, David. 1989. “How social comparison mediates the relation between ability grouping practices and students achievement expectancies in mathematics.” *Journal of Educational Psychology* 81(2): 178‑189.

Richer, Stephen 1976. "Reference-group theory and ability grouping: A convergence of sociological theory and educational research." *Sociology of Education* 49(1): 65‑71.

Rist, Ray. 1970. "Student Social Class and Teacher Expectations: The Self-Fulfilling Prophecy in Ghetto Education." *Harvard Educational Review* 40(3): 411-451.

Rist, Ray. 1973. *The urban school: A factory for failure*. Cambridge, MA: MIT Press.

Rosenbaum, James E. 1976. *Making inequality: the hidden curriculum of high school tracking.* New York: Wiley.

Rosenbaum, Paul and Donald Rubin. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika* 70(1): 41-55.

Sakura-Lemessy, Ian, Katrina Carter-Tellison, Donna-May Sakura-Lemessy. 2009. “Curriculum placement, race, class, and gender differences in wage growth amongst non-college workers: Evidence from the NELS 1988-2000 data.” *Journal of African American Studies* 13(4): 406-430.

Schumm, Jeanne, Sally Moody, and Sharon Vaughn. 2000. “Grouping for reading instruction: Does one size fits all?” *Journal of Learning Disabilities* 33(5): 477–488.

Schwartz, Frances. 1981. "Supporting or Subverting Learning: Peer Group Patterns in Four Tracked Schools." *Anthropology and Education Quarterly* 12(2): 99-121.

Slavin, Robert. 1987. "Ability Grouping and in Elementary Schools: A Best-Evidence Synthesis." *Review of Educational Research* 57(3): 293-336.

Slavin, Robert. 1990. "Achievement effects of ability grouping in secondary schools: A best-evidence synthesis." *Review of Educational Research*, 60(3): 471-499.

Smith, Alison, Lee Jussim, Jacquelynn Eccles, Michelle VanNoy, Stephanie Madon, and Polly Palumbo. 1998. "Self-Fulfilling Prophecies, Perceptual Biases, and Accuracy at the Individual and Group Levels." *Journal of Experimental Psychology* 34: 530-561.

Sørenson, Aage B. 1970. "The Organizational Differentiation of Students and Educational Opportunity." *Sociology of Education* 43(4): 355-76.

Stanovich, Keith. 1986. "Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy" *Reading Research Quarterly* 21(4): 360–407.

Stipek, Deborah. 2002. *Motivation to Learn: From Theory to Practice (4th Edition).* Needham Heights, MA: Allyn and Bacon.

Strike, Kenneth. 1983. "Fairness and Ability Grouping." *Educational Theory* 33(3 & 4):125-34.

Tach, Laurie, and George Farkas. 2006. "Learning-related behaviors, cognitive skills, and ability grouping when schooling begins." *Social Science Research* 35(4): 1048–1079.

Thompson, James. 1967. *Organizations in Action: Social Science Bases of Administrative Theory.* New Brunswick: Transaction Publishers.

Trautwein, Ulrich, Oliver Lüdtke, Herber Marsh, Olaf Köller, and Jürgen Baumert. 2006. "Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics." *Journal of Educational Psychology* 98(4): 788-806.

Tyack, David. 1974. *The One Best System: A History of American Urban Education*. Cambridge, Massachusettes: Harvard University Press.

Tyson, K. 2011. *Integration interrupted: Tracking, Black students, and acting White after Brown.* New York, NY: Oxford University Press.

Useem, Elizabeth. 1992. " Getting on the Fast Track in Mathematics: School Organizational Influences on Math Track Assignment." *American Journal of Education* 100(3): 325-353.

Weinstein, Rhona. 1976. “Reading group membership in first grade: Teacher behaviors and pupil experience over time.” *Journal of Educational Psychology*, 68(1): 103-116.

Whipple, Guy. 1919. *Classes for Gifted Children*. Bloomington, IL: Public School Publishing.

Yonezawa, Susan and Makeba Jones. 2006. “Students’ Perceptions of Tracking and Detracking.” *Theory into Practice* 45(1): 15-23.

Yonezawa, Susan, Amy Wells, and Irene Serna 2002. "Choosing tracks: 'freedom of choice' in detracking schools." *American Education Research Journal* 39(1): 37-68.

Zimmer, Ron. 2003. "A new twist in the educational tracking debate." *Economics of Education Review* 22(3): 307-315.