STRATIFICATION, SKILL GROUPING, AND LEARNING
TO READ IN FIRST GRADE

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By
Dennis J. Condron, M.A.

* * * * *

The Ohio State University
2005

Dissertation Committee:

Professor Douglas B. Downey, Advisor
Professor Vincent J. Roscigno
Professor Zhenchao Qian

Approved by

__________________________
Advisor
Department of Sociology
ABSTRACT

Curriculum differentiation – providing different instruction to groups of students based on their skills or perceived abilities – is a practice that affects a vast majority of students in the U.S. For sociologists, the role of curriculum differentiation in unequal educational opportunities and outcomes is directly relevant to theoretical perspectives on education and stratification. Do schools reduce inequality, or do they reproduce or even exacerbate inequality?

In this theoretical context, I focus on one common type of curriculum differentiation: Within-classroom skill grouping for reading instruction during elementary school. Analyzing first-grade data from the Early Childhood Longitudinal Study-Kindergarten Cohort, I find evidence that socioeconomically disadvantaged, Black, and Hispanic students tend to be placed into lower-ranked reading groups more often than their advantaged counterparts, primarily due to disparities in skills that are already present when school begins. Improving upon past research, I compare the reading gains of students placed into differentially-ranked groups to those of similar students whose teachers do not use skill grouping. This analytic strategy reduces the likelihood that characteristics of students other than their group placement bias the estimates of the effects of group placement on learning. Findings suggest that high-grouped students gain more, and low-grouped students gain less, than similar students in
non-grouped classrooms. In addition, within-student analyses across kindergarten and first grade suggest that students learn more the year they are placed into a high-ranked group, and less the year they are placed into a low-ranked group, compared to the year in which their teacher did not use skill grouping. Finally, I assess the role of skill grouping in the early emergence of socioeconomic and racial gaps in learning. Evidence that skill grouping is to blame for such gaps is scant, but limitations of the data make it difficult to draw conclusions.

I discuss the implications of these findings for theoretical perspectives on education and stratification and for contemporary policies aimed at reducing gaps in learning. For scholars of educational stratification, as well as parents, policy makers, and educators concerned about inequality, skill grouping within elementary classrooms should rank high among school-related practices considered to promote disparities in skills.
ACKNOWLEDGMENTS

I first want to thank two individuals in particular who have had a profound influence on my professional development in graduate school, which culminates in the completion of this dissertation: Professors Vincent Roscigno and Douglas Downey. They have been exceptional mentors in countless ways throughout my time at Ohio State, and have taught me invaluable lessons about Sociology and being an academic sociologist. Doug’s insistence on conducting research of the highest quality and his knack for staying focused on big-picture questions have had a tremendous impact on me as I have formed my own approach to doing research. Vinnie’s guidance has consistently gone far above and beyond the call of duty, and I am very fortunate to have learned a great deal from him about so many things.

I also want to thank Professor Zhenchao Qian for his helpful role on my candidacy exam and dissertation committees. Professors Robert Kaufman, Randy Hodson, David Jacobs, and Paul Bellair, along with Dr. Paul von Hippel, all provided insights regarding statistical methods for which I am grateful. Professors Richard Lundman, James Moody, Liana Sayer, and Andrew Martin offered constructive feedback at various stages of this project. Among my graduate student colleagues, I am especially grateful for the intellectual and personal friendships of Marc Dixon, James (Jim) Sutton, Daniel Tope, Donna Bobbitt-Zeher, and Clayton Peoples.
On a more personal level, I want to express my gratitude for the boundless and unconditional support of Susanne Condron. Spouses are “supposed to” provide emotional support and encouragement when the going gets tough, but Susanne has done so much more. Debating issues of stratification and education, listening to practice presentations, sharing syntax, and helping with Excel charts are just a few examples of her more tangible contributions to the completion of this dissertation. I truly have it made.

Finally, this research was financially supported by a Presidential Fellowship from the Graduate School of the Ohio State University.
VITA

December 25, 1976 . . . . . . . . . . . . Born – Southfield, Michigan

1999 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . B.A., Sociology, University of Michigan-Flint

1999-2001 . . . . . . . . . . . . . . . . . . . . . . Graduate Teaching Associate
Ohio State University

2001 . . . . . . . . . . . . . . . . . . . . . . . . . . . M.A., Sociology, Ohio State University

2001-2004 . . . . . . . . . . . . . . . . . . . . . . Graduate Teaching and Research Associate
Ohio State University

2004-2005 . . . . . . . . . . . . . . . . . . . . . . Presidential Fellow, Ohio State University

PUBLICATIONS


FIELDS OF STUDY

Major Field: Sociology
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>Vita</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>ix</td>
</tr>
<tr>
<td>List of Figures</td>
<td>x</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2. Theory and Research on Stratification and Curriculum Differentiation</td>
<td>10</td>
</tr>
<tr>
<td>3. Data and Methods</td>
<td>36</td>
</tr>
<tr>
<td>4. Disparities in Skill Group Placement</td>
<td>48</td>
</tr>
<tr>
<td>5. The Impact of Skill Grouping on Reading Gains</td>
<td>55</td>
</tr>
<tr>
<td>6. Inequality, Socioeconomic Status, and Race</td>
<td>67</td>
</tr>
<tr>
<td>Inequality in the Test-Score Distributions of Grouped and Non-Grouped Students</td>
<td>68</td>
</tr>
<tr>
<td>Socioeconomic and Racial Disparities in Learning: Does Grouping Matter?</td>
<td>73</td>
</tr>
<tr>
<td>Theoretical Implications</td>
<td>79</td>
</tr>
<tr>
<td>7. Discussion and Conclusion</td>
<td>81</td>
</tr>
<tr>
<td>Limitations and Implications for Future Research</td>
<td>87</td>
</tr>
<tr>
<td>Conclusion</td>
<td>92</td>
</tr>
<tr>
<td>List of References</td>
<td>94</td>
</tr>
<tr>
<td>Appendices:</td>
<td></td>
</tr>
<tr>
<td>Appendix A: Tables and Figures</td>
<td>101</td>
</tr>
<tr>
<td>Appendix B: Measuring and Modeling Learning</td>
<td>119</td>
</tr>
<tr>
<td>Appendix C: Notes on Robustness of Results</td>
<td>128</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Descriptive Statistics for Measures Used in the Study</td>
<td>102</td>
</tr>
<tr>
<td>2</td>
<td>Generalized Least Squares Regression Coefficients for Hypothesized Predictors of Skill Group Rank</td>
<td>104</td>
</tr>
<tr>
<td>3</td>
<td>Results of T-Tests Comparing Means of Selected Characteristics of Grouped and Non-grouped Students’ Schools, Classrooms, and Teachers</td>
<td>105</td>
</tr>
<tr>
<td>4</td>
<td>Generalized Least Squares Regression Coefficients Estimating the Effects of Skill Group Placement on Reading Gains</td>
<td>106</td>
</tr>
<tr>
<td>5</td>
<td>Results of Cross-Sectional Time-Series Analyses with Fixed Effects Estimating Within-Student Reading Gains Across Kindergarten and First Grade</td>
<td>108</td>
</tr>
<tr>
<td>6</td>
<td>Growth in Socioeconomic and Racial Gaps in Reading Skills Among Grouped and Non-Grouped Students</td>
<td>109</td>
</tr>
<tr>
<td>7</td>
<td>Results of Generalized Least Squares Regressions Testing for Mediation of Socioeconomic Gaps in Reading Gains</td>
<td>110</td>
</tr>
<tr>
<td>8</td>
<td>Results of Generalized Least Squares Regressions Testing for Mediation of Racial Gaps in Reading Gains</td>
<td>111</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hypothesized Effects of Skill Grouping on Reading Gains</td>
<td>112</td>
</tr>
<tr>
<td>2</td>
<td>Socioeconomic Disparities in Skill Group Placement</td>
<td>113</td>
</tr>
<tr>
<td>3</td>
<td>Racial Disparities in Skill Group Placement</td>
<td>114</td>
</tr>
<tr>
<td>4</td>
<td>Variance in Reading Skills in the Fall and Spring of First Grade for Grouped and Non-grouped students</td>
<td>115</td>
</tr>
<tr>
<td>5</td>
<td>Change in Variance in Reading Skills from Fall to Spring by Skill Group Status</td>
<td>116</td>
</tr>
<tr>
<td>6</td>
<td>Scatter Plot of Fall and Spring Reading IRT Scores</td>
<td>117</td>
</tr>
<tr>
<td>7</td>
<td>Scatter Plot of Fall Reading IRT Score and Gains During First Grade</td>
<td>118</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

Why are some children “smarter” than others? Many factors play a role in shaping students’ learning, some of which are rooted in biological differences (“nature”) while others are rooted in environmental influences (“nurture”). Some students, for example, might have developmental or learning disabilities present at birth that prevent them from learning at the same level as students who are not afflicted with such disadvantages. Beyond extreme cases of disabilities or exceptional intelligence, there is also a great deal of debate over the extent to which everyday differences in educational performance are genetic, or “built-in.”

The controversy surrounding Herrnstein and Murray’s (1994) *The Bell Curve* exemplifies this debate. Herrnstein and Murray (1994) emphasized the role of heredity in shaping individuals’ (and socioeconomic and racial groups’) intellectual capacities. In contrast, Fischer and colleagues’ (1996) *Inequality by Design* countered that inequality in cognitive skills is driven mainly by inequality in the social and environmental circumstances of individuals and groups.

In general, sociological approaches to understanding why some people are seemingly smarter than others emphasize the role of the many social forces that influence
us in various ways every day. For example, the social class position of the family into which a child is born shapes the extent to which supplemental educational resources such as parental involvement, books, computers, and tutors are available (Gamoran 2001; Lareau 1987). Moreover, residential segregation by race and social class interacts with an unequal system of school funding to result in an education system that offers a lower quality of education to poor and racial minority families. Schools serving White and better-off students tend to be new and well-equipped with the latest technology and advanced curricula, while schools serving poor and minority students are often aging, falling apart, and lacking important resources needed to provide the same opportunities for children to learn (Kozol 1992).

There is little doubt that such variations in children’s social, economic, and educational environments matter, and this notion has influenced policy decisions. For example, many education policies over the last few decades have been based on the assumption that altering environmental factors can affect students’ learning. Head Start, reducing class sizes, and increasing funding for disadvantaged schools are all ways in which policy makers have implemented changes in the educational landscape designed to close achievement gaps. A contemporary example is the federal “No Child Left Behind” Act, which calls for educators to be held accountable for students’ mastery of skills and allows parents to remove their children from ineffective schools in favor of a more positive learning environment. Clearly, contextual factors such as teachers and schools are thought to matter in important ways.
This study aims to help us better understand how one important social aspect of schooling might lead some students to learn more than others – the within-classroom grouping of similarly-skilled students for reading instruction. When school begins, teachers are faced with the dilemma of providing instruction to a classroom full of students whose skills and rates of learning vary dramatically. Teachers may opt to instruct all students together by moving through the material at a pace thought to be appropriate for *most* of the students in the class. However, in this scenario, the faster learners might not be challenged by the pace and difficulty of instruction, while, at the same time, the slower learners might be unable to keep up. A common way in which teachers respond to this challenge of “managing heterogeneity” (Berliner and Biddle 1995:321) in students’ reading skills is by conducting lessons with students placed into small groups. The students in a given group are similar in terms of perceived skill levels or capacity to learn, and teachers spend a period of time working with each group separately. By working with small groups of students that are similarly skilled, teachers are able to proceed through the material at a pace and level of difficulty that is appropriate for *all* students in the classroom. In other words, teachers can cover more material (perhaps even using more advanced books) in the groups with highly-skilled students and less material in the groups with lower-skilled students as deemed appropriate.

The goal of homogenous skill grouping (often called “ability grouping”) is to increase the effectiveness of instruction (Gamoran 1989) such that grouping “contributes positively to the academic progress of all students” (Kerckhoff 1986:842). If skill
grouping works as it is intended to, all students should benefit from small-group instruction relative to the full-classroom instructional setting. When grouped, the faster learners are not held back by a pace of instruction that fails to sufficiently challenge them; at the same time, the slower learners are not left behind by a pace of instruction that is too challenging for them. Whether a student is placed into a high- or low-ranked group, that student should learn more than she or he would have if she or he had not been placed into a homogeneous group at all. This view, which I refer to as the equality of opportunity perspective, suggests that skill grouping is a rational and efficient organizational method of managing heterogeneity in students’ skills that has “pedagogical advantages over heterogeneous grouping” (Hallinan 2003:124).

Another view, however, suggests that skill grouping within classrooms does not benefit all students, but instead exacerbates pre-existing disparities in academic skills. Many have argued that homogeneous skill grouping tends to segregate students according to socioeconomic status (SES) and race within schools and/or classrooms; studies have consistently found that poor and minority students tend to be overrepresented in lower-ranked groups (Oakes, Gamoran, and Page 1992; Slavin 1987). There is also evidence suggesting that students in higher-ranked groups learn more than students in lower-ranked groups, net of social background characteristics and prior measures of academic skills (Kerckhoff 1986; Pallas et al. 1994). These findings have led some to conclude that skill grouping widens the inequality in academic performance with which students begin school and is thus an institutional mechanism through which
the education system contributes to inequality (Gamoran 1992; Gamoran et al. 1995). I refer to this view as the reproduction of inequality perspective.

How can we reconcile these contrasting perspectives? Does skill grouping benefit all students, or only some students, possibly at the expense of others? Surprisingly, researchers have been unable to generate evidence solid enough to allow them to reach confident conclusions. As I discuss in the next chapter, our understanding of this educational practice and its broader implications for inequality in cognitive skills is limited due to a lack of high-quality studies.

The lack of good research is unfortunate because skill grouping is widely used and affects a vast majority of young students in the U.S. Drawing from an earlier study, Entwisle and Alexander (1993) reported that skill grouping was found in 90% of elementary schools in the U.S. Similarly, in their Baltimore study, only one out of the 20 schools originally sampled did not use within-classroom skill grouping (Pallas et al. 1994). A recent review of the literature reported that “[a]bility grouping for reading instruction appears nearly universal, especially in the early grades” (Loveless 1998) while Ferguson (1998:325) noted that “[a]lthough the practice is not documented by any national survey, it seems that the vast majority of elementary schools in the United States use in-class ability grouping for reading and whole class instruction for other subjects.”

Given the lack of high-quality studies and the widespread use of skill grouping, more research is clearly needed.

This study offers several important contributions to the literature. First, it focuses on young children and the within-classroom contexts that influence their learning as their
educational trajectories are launched early on. Most studies of educational grouping focus on between-classroom tracking among older students. Second, the analyses use a recent, high-quality, nationally representative data source, meaning that the results provide an overall picture of what is occurring nationally when it comes to within-classroom grouping for reading instruction. Most studies of this nature draw on only a handful of classrooms sampled from within a single district or even a single school. Third, since students in the sample were studied in both the fall and spring and tested using advanced scoring techniques, the data allow me to measure and predict the gains in reading skills that students make over the course of the school year. This is an important advantage over most studies, which tend to rely on report card marks or other evaluations that are not well-suited for measuring gains in skills over time. A fourth contribution is that my analyses compare the learning of (a) similar grouped and non-grouped students and (b) the same student in grouped and non-grouped years of schooling. These analytic approaches are better able to rule out characteristics of children other than their skill group placements as explanations for why students placed into different groups have different learning outcomes.

My analyses seek to understand the patterns, processes, and potential inequalities related to early elementary within-classroom skill grouping. After reviewing the literature on stratification and curriculum differentiation (Chapter 2), I address the question of whether poor and minority students tend to be placed into lower-ranked groups more frequently than their more advantaged classmates, and whether teachers base their grouping decisions on such ascribed characteristics or instead on students’
level of academic preparedness (Chapter 4). I then provide a rigorous test for effects of group placement on learning, analyzing (a) whether grouped students learn more than non-grouped students as a whole and (b) whether students placed into low-, middle-, and high-skill reading groups learn more or less than similar students whose teachers do not group (Chapter 5). This is followed by analyses of whether skill grouping produces a more unequal distribution of learning outcomes relative to full classroom instruction, and whether skill grouping is partially to blame for test-score gaps by race and socioeconomic status (Chapter 6). Results of these empirical analyses suggest that socioeconomically disadvantaged, Black, and Hispanic students tend to be placed into lower-ranked reading groups than their advantaged counterparts, and that skill grouping promotes unequal learning outcomes when compared to non-grouped instruction. I discuss the implications of these findings for theoretical perspectives on education and stratification, and for policy aimed at closing gaps in educational achievement and attainment.

**Clarification of Terms**

Before continuing, it is necessary to clarify the meanings of some of the terms commonly used in this study. Generally speaking, the process of skill grouping analyzed here is a type of curriculum differentiation – making “different knowledge available to different groups of students” (Oakes et al. 1992:570). This practice takes many different forms and occurs at virtually all stages of the educational process. There are also several terms used to refer to the different types of curriculum differentiation, and confusion can result from the use of these terms interchangeably.
“Curriculum differentiation,” as noted above, is the broadest term representing the concept of providing different material to different students. Traditionally, “tracking” has been used to describe the process whereby secondary school students are channeled into different classes with varying levels of difficulty (i.e., one ninth-grader is assigned to take remedial algebra while another is assigned to general algebra while yet another may be assigned to a geometry class with mostly tenth-graders). In contrast, “ability grouping” has typically referred to the separation of elementary school students into homogenous groups within the same classroom (see Gamoran 1992).

Complicating the issue, however, is the notion that traditional, more rigid systems of high-school tracking – whereby students were typically tracked high or low in all subject areas – have been largely dismantled in recent decades. Instead, as Lucas (1999) pointed out, students today often take courses of varying levels of difficulty in different subject areas. This has led some to apply the term “ability grouping” to between-classroom separation at the secondary level as well (i.e., Hallinan 2003; Kerckhoff 1986). In other cases, researchers use “tracking” and “ability grouping” somewhat interchangeably (i.e., Hoffer 1992). Tracking and ability grouping are similar in that they both aim to provide a different level of instruction to different students; however, placing young students into small, homogeneous groups within classrooms for a short lesson is not the same as separating older students into different classrooms entirely. While much of this might appear to be a matter of preference, researchers should, nonetheless, at least specify the assumptions that lead them to use one term or another.
In this study, I use the term “skill grouping” rather than the more common “ability grouping” to refer to the separation of students into small homogeneous groups within classrooms for reading lessons. “Grouping” seems more appropriate for describing curriculum differentiation within classrooms than between classrooms (where “course taking” might be a better term). The use of the word “skill” is another key distinction. “Ability” suggests an innate, fixed attribute of students while “skill” more accurately reflects the different rates of learning on which teachers’ grouping decisions are based. This is consistent with the approach of Mosteller, Light, and Sachs (1996:799), who preferred “‘skill grouping’ rather than ‘ability grouping’ because the latter suggests a sense of permanence in a quality that we believe might be modified by education, training, and practice. Skill grouping, on the other hand, suggests that students sharing a similar current skill level are grouped together for purposes of instruction.” When quoting other authors, I retain their use of the term “ability grouping.”
CHAPTER 2

THEORY AND RESEARCH ON STRATIFICATION AND CURRICULUM DIFFERENTIATION

Over the past few decades, the role of education in the process of stratification has emerged as arguably the predominant overarching issue in the sociology of education. Do schools help to equalize opportunities among individuals from different social origins, or do they reproduce the existing stratification hierarchy? As Hurn (1993:161) put it, “[d]o schools merely respond in neutral or objective fashion to the differences that students bring with them to school, or do they play a part in magnifying or even creating these differences?”

The effect of skill grouping on students speaks directly to this broader theoretical debate because it is an organizational practice occurring within schools that is implemented with the specific purpose of helping teachers manage “the differences that students bring with them to school” – in this case, differences in academic skills. If skill grouping is a neutral response to variation in students’ skills that benefits all students, then it is likely to promote equal opportunity. If skill grouping is a non-neutral response that benefits only some students, then it is likely to contribute to inequality.
In this chapter, I begin by discussing the role of educational opportunities and outcomes in the process of stratification from these two broad theoretical traditions. I then review and build upon the existing literature on curriculum differentiation, with a particular emphasis on studies of within-classroom skill grouping in elementary school. I conclude by outlining my specific research questions and the expected findings predicted by each theoretical perspective.

Two Perspectives on Education, Stratification, and Skill Grouping

Two general theoretical orientations have dominated discourse in the sociology of education and stratification, and these contrasting views have also shaped research on curriculum differentiation.

Education and Equality of Opportunity

In the 19th century, Horace Mann, known as the “father of public education,” declared that schools are the “great equalizer of the conditions of men” (Cremin 1957:87). A key assumption among those promoting the idea of a common school system was that mass education would allow all people, no matter what their social origin, an equal opportunity to compete for occupational and economic success and to participate in the political functioning of society (Ravitch 2001). Beyond its role in the policy arena early on, the idea that schools are the great equalizer, along with the assumption that education is a meritocratic process, surfaced in sociological theory and research throughout the twentieth century and is still present to this day. Becker (1964), for instance, argued that educational attainment is an investment in “human capital” which improves individuals’
knowledge and skills and, ultimately, increases earnings. Similar assumptions guided status attainment research, which tended to argue that an individual’s occupational status is driven by one’s own educational achievements more so than one’s social origin (Blau and Duncan 1967).

There is a solid body of empirical evidence in favor of the view that schools are the great equalizer. As Spring (2001) pointed out, increasing access to educational opportunities in the hundred years following slavery allowed African Americans to increase their literacy rate from a mere 7% during the Civil War to about 90% by 1950. Furthermore, the Black/White test score gap has continued to narrow over time, and schools appear to have played a positive role in that process (Jencks and Phillips 1998). Evidence from the National Assessment of Educational Progress (NAEP, known as “the nation’s report card”) has revealed that Black and Hispanic students have gained ground on White students over the past few decades (Berliner and Biddle 1995). These findings all lend support to the notion that schools provide disadvantaged students with an arena in which they can improve their educational and occupational prospects.

Another line of research also supports this position, implicating the out-of-school environment as the primary driving force behind the class and racial gaps in educational achievement that persist. Studies have attempted to separate the effects of school and non-school factors on learning by comparing students’ learning during the school year to that which occurs during the summer months – when school is not in session. The idea is that since schools do not shape learning when students are not attending, gains or losses in skills that occur during the summer months are mainly attributable to variations in
non-school factors such as family socioeconomic status, neighborhood conditions, and the like. With measures of skills collected during both the fall and spring of multiple school years, researchers can exclude summer learning and the effects of non-school factors when looking at the effects of schools. “...[T]hese effects are disguised when rates of growth are computed annually, as they usually are, because complex interactions between in-school and out-of-school learning, home background, and minority status are obscured” (Entwisle and Alexander 1992:83). Once summer gains or losses are measured and separated from those occurring during the school year, it is possible to make fairly straightforward comparisons of summer and school-year learning.

Findings from this line of inquiry tend to suggest that schools reduce inequality. One study of several hundred youngsters in Baltimore from 1982 to 1984, for instance, found that “... home disadvantages are compensated for in winter because, when school is in session, poor children and better-off children perform at almost the same level ... It is mainly when school is not in session that consistent losses occur for poorer children” (Entwisle and Alexander 1992:82). More recent analyses, using the same data I use in this study, led to a similar conclusion (Downey, von Hippel, and Broh 2004). These studies suggest that there would be more inequality in cognitive skills between low- and high-SES students if they did not attend school at all; thus schools are considered to have an equalizing effect on disparities in skills by SES. Alexander (1997) attributed the narrowing of test score gaps between advantaged and disadvantaged children to improvements in public education, such as increased spending, smaller class sizes, and more access to advanced curricula. It has been more difficult to absolve schools of
contributing to racial gaps in skills, as Entwisle and Alexander (1992) reported no racial differences in summer gains/losses once SES was controlled, and Downey et al. (2004) found that schools reduce socioeconomic inequality in skills but not racial inequality.

*Education and the Reproduction of Inequality*

As the above discussion illustrates, there are both theoretical and empirical reasons to believe that education is a mechanism of equal opportunity. However, another view of education and stratification comes to a different conclusion. In contrast to the view that schools are the great equalizer, scholars in the conflict theoretical tradition have suggested that the education system instead reproduces the existing stratification hierarchy. Bowles and Gintis (1976), for example, argued from a Marxist perspective that schools reinforce the class inequality inherent in the capitalist mode of production. Sons and daughters of the elite attend high-status private schools that channel them into high-status positions in the stratification hierarchy, while most students attend schools that promote the attitudes and values demanded by employers – docility, compliance, and punctuality. At the same time, the credential system in the U.S. is thought to reproduce inequality because higher-status people have an advantage in obtaining the levels of education required for high-status jobs and economic success (Collins 1979). In these ways, the education system reinforces the existing stratification hierarchy despite the rhetoric of equal opportunity.

More specifically, there are processes occurring within schools that hold important implications for the reproduction of inequality position. First, students of different social origins attend elementary and secondary schools in the U.S. that are, to
quite an extent, separate and unequal (Coleman et al. 1966; Kozol 1992). The class and racial segregation of schools has persisted, due largely to continuing high levels of residential segregation (Massey and Denton 1993; Rivkin 1994). In recent times, school segregation has become even more pronounced, due to a “quiet reversal of Brown v. Board of Education” following a brief era of limited racial integration (Orfield et al. 1996). Related to these patterns of segregation is the reality that schools in the U.S. are unequally funded and offer students drastically unequal opportunities to learn. Kozol (1992), for instance, demonstrated that the schools of poor and minority children in America often lack the variety of course offerings and advanced placement courses that higher-status students are able to use in preparing for college and eventual occupational success. These disparities manifest themselves in depressed levels of academic achievement in minority segregated schools and call into question the view that education provides equal opportunity for minority students (Bankston and Caldas 1996; Roscigno 1998).

Second, there has been concern that the education system unequally rewards cultural capital, or cultural “cues” and “competencies” that help students “negotiate their educational experience” (Lamont and Lareau 1988:155). One line of inquiry focuses on social class and the ways in which school practices reward the cultural capital of higher-status students (Bourdieu 1977). Since lower-status students often lack the cultural capital needed for success in school, such as the degree of parental involvement expected by schools/teachers (Lareau 1987), the education system is thought to favor students from higher social classes. There is some empirical evidence to support these assertions,
as DiMaggio (1982) found that students with greater participation in cultural events (i.e., literary, artistic, and musical activities such as going to symphony concerts) had better grades in high school than their peers with less cultural capital.

A more specific approach has considered the possibility that there is a cultural mismatch between students and teachers of different social origins, which contributes to the educational struggles of disadvantaged students (Downey and Pribesh 2004). One study, for example, found that “high-status teachers, both black and white, experience special difficulties relating to minority youngsters” and concluded that these difficulties likely influence early racial gaps in academic achievement (Alexander, Entwisle, and Thompson 1987:679). Similarly, Roscigno and Ainsworth-Darnell (1999) found evidence that the effects of cultural capital on high-school achievement are weaker among lower-SES and racial minority students, suggesting that teachers differentially reward the cultural capital of students from different social backgrounds in a way that places poor and minority students at a disadvantage in terms of achievement.

Reconciling the Two Perspectives

As the above discussion illustrates, there is evidence supporting both the equality of opportunity and reproduction of inequality perspectives. These patterns clearly must be reconciled if theory and research on education and stratification are to move forward. How might it be that schools lessen and reproduce inequality at the same time?

There are two relevant points to be made. First, it depends on the comparison one is making. The research on seasonal differences in learning suggests that inequality in cognitive development is greater in the absence of schooling than in its presence. The
key comparison in this line of research is unequal learning during the school year relative to that occurring over the summer. Since there would be more inequality if schools were abolished (as a hypothetical example), schools are attributed a positive role in the explanation of unequal learning outcomes. Analyzing unequal learning across educational contexts during the school year, however, is a different comparison that focuses on processes occurring within schools that could potentially be problematic but not captured in a summer/school-year comparison. While inequality in cognitive skills might be reduced during the school year relative to the summer, there is still significant inequality in both opportunities and outcomes when school is in session. That is, even if there are differential gains/losses during the summer compared to the school year that suggest a positive role of schooling (consistent with the equality of opportunity perspective), there are also disparate gains/losses that are exclusive to the school year and might be better understood from the reproduction of inequality perspective.

A second issue concerns the relative extent of inequality in children’s school and non-school environments. Both undoubtedly vary considerably, but it is possible that there is less inequality in the realm of schooling than there is in other areas of children’s lives. If this is true, it would follow that the most disadvantaged children in society benefit greatly from schooling because going to school is an improvement over their out-of-school environment (i.e., there are resources for learning at school compared to few or none at home, there is free lunch at school compared to no food at home, etc.). At the same time, the most advantaged children benefit less from schooling because going to school is not much of an improvement over their non-school environment (i.e., they
already have resources for learning at home, there is abundant nutrition at home, etc.).

“As a result, a disadvantaged child attending a low-quality school can enjoy a larger ‘school boost’ than an advantaged child attending a high-quality school” (Downey et al. 2004:614).

Under this line of thinking, it is entirely possible that schools promote equal opportunity by reducing the inequality that emerges in their absence while they simultaneously reproduce inequality by conferring advantages and disadvantages on different kinds of students when school is in session. Because it is implemented on the basis of “differences that students bring with them to school” (Hurn 1993:161), a sound understanding of the effects of skill grouping on students’ learning outcomes is crucial to these theoretical debates over the role of education in the broader system of stratification.

**Extending Prior Research on Skill Grouping**

What do we know about homogeneous skill grouping within elementary classrooms, and how can additional research improve sociological theory and inform parents’, teachers’, and policy makers’ understanding of its effects on students? As noted above, there has been a lack of studies of early elementary skill grouping in recent times, at least relative to the proliferation of analyses of high-school tracking (i.e., Hallinan 1996; Gamoran and Mare 1989; Lucas 1999; Lucas and Good 2001; Oakes 1985). The existing studies, however, have tended to focus on two central issues – the process through which students are placed into groups and the subsequent effect of group placement on students’ learning. A third issue, which has received little or no attention, is whether there are systematic differences between grouped and non-grouped students’ teachers, classrooms,
or schools. Since most studies have focused on grouped students only (a point I note below), we know little about what these differences might be. Capturing them is important in this study in order to ensure that potential effects of skill grouping on learning are not really due to unmeasured variations in the learning environments of grouped and non-grouped students.

**Group Placement**

The primary question surrounding placement into skill groups has been whether it is a meritocratic process. In other words, are students placed into groups based on their effort and level of academic skills, or do ascribed characteristics such as socioeconomic status and race shape teachers’ judgements?

In Rist’s (1970) classic study, students’ visible social class characteristics influenced the teacher’s perception of their skills, and lower-class children were disproportionately placed into the lower-ranked reading groups. This study raised a great deal of concern over whether teachers sometimes discriminate against poor and/or minority students when creating reading groups, but more recent quantitative work has not supported Rist’s (1970) findings. Studies have found that once measures of students’ skills are taken into account, any initial relationship between ascribed characteristics and group rank tends to weaken or dissolve completely (Haller 1985; Haller and Davis 1980, 1981; Gamoran 1986, 1989; Pallas et al. 1994). In Pallas et al.’s (1994) study of Baltimore youngsters, for instance, initial models suggested a slight reading group advantage for children of more highly educated parents, but this advantage was explained by higher test scores and grades among those students. The overall conclusion from this
body of research is that skill group placement appears to be based not on ascribed characteristics, but rather differences in students’ skills, which are often associated with ascribed characteristics. Since disadvantaged students typically begin the school year with fewer skills, they frequently become overrepresented in lower skill groups.

Studies have also found that teachers take more than students’ skills into account when forming groups. For example, Hallinan and Sorensen (1983) and Eder (1983) pointed out that teachers often predetermine the number and size of the groups they will create. Because teachers have limited time, they frequently put a cap on the number of groups they will form, allowing sufficient time to meet with each group. In addition, group size is often shaped by teachers’ desires to keep the instructional climate manageable in terms of students’ attention and behavior. The teacher in Eder’s (1983) study, for instance, reported an unwillingness to have groups with any more than seven students. These studies call our attention to the reality that factors other than academic skills and ascribed characteristics may also influence group placement (see also Hallinan and Sorensen 1986).

Despite the evidence suggesting that group placement is meritocratic – and also less than perfectly related to students’ cognitive skills – many studies of curriculum differentiation at both the elementary and secondary level have found a pattern of disproportionate lower group and track placement among poor and minority students (Oakes et al. 1992; Slavin 1987). In other words, even though teachers appear to base their grouping decisions on students’ skills, poor and minority students are more likely to be placed into lower-ranked groups because they enter school with fewer skills as a result
of disadvantages in their non-school environment (Entwisle and Alexander 1993).

Ultimately, then, the concern is that skill grouping contributes to socioeconomic and racial segregation within schools. To consider the implications of this issue, I not only predict students’ group placements net of skills and a variety of other attributes, but I also test for bivariate racial and socioeconomic differences in group rank in order to establish whether there is a pattern of disproportionate low group placement among disadvantaged students. If so, this pattern becomes important when considering the potential effects of group placement on learning, an issue I discuss below.

*Differences between Grouped and Non-Grouped Students’ Learning Environments*

Researchers studying only students in grouped classrooms have not had to contend with the question of why some teachers group and others do not. In the sample of first graders I analyze in this study, the teachers of roughly 30% of the students reported that they do not use achievement-based groups for reading lessons. What might be some of the ways in which the teachers, classrooms, and schools of grouped and non-grouped students differ?

The literature on skill grouping offers few insights into this issue, likely because, as noted above, grouping is so widely used. The presence of a large number of non-grouped students in the nationally representative data used in this study suggests a possible decline in the use of skill grouping over time, which perhaps has resulted from criticism of it.

One possibility is that grouping occurs more frequently in schools serving disadvantaged students due to a desire to “save a few” of the best and brightest students
by exposing them to accelerated curricula (Oakes et al. 1992). It seems reasonable that students in larger schools and classrooms may be grouped more frequently than students in smaller schools and classrooms, since there may be more heterogeneity in students’ skills in larger settings (Loveless 1998). Another possibility is that students are grouped more often in public schools than in private schools, given the likelihood that public school students’ skill levels and social backgrounds are more heterogeneous than those of private school students (Lee and Bryk 1989). I also consider whether the race, age, level of education, and years of experience of the teacher are associated with using skill grouping. It could be that older, more experienced, and more highly educated teachers are less likely to group if, for example, they have a better handle on teaching or are more in tune with the criticisms of skill grouping relative to their younger, less experienced, and less educated counterparts. Finally, students in classrooms with more time and space constraints (Eder 1983; Hallinan and Sorensen 1983) or disorganization may be more likely to be grouped – i.e., if the students’ behavior is disruptive or if the teacher lacks textbooks or the classroom space needed to conduct lessons in large groups.

Effects of Skill Group Placement on Learning

Does the group into which a student is placed influence how much he or she will learn? This straightforward question has received considerable attention; yet, due to limitations of prior research, the literature has not yielded a straightforward answer. As Mosteller et al. (1996:812) put it, “[a]fter examining fifteen experiments involving skill grouping, we find little evidence that skill grouping has a major impact, either positive or negative, on students’ cognitive learning.”
There are three main reasons why there is not a definitive conclusion as to whether skill grouping for reading influences students’ learning. The first stems from the fact that there are several different forms of curriculum differentiation – within-classroom, between-classroom, between-classroom only for a certain subject, grouping in elementary schools, tracking in high schools, etc. (see Oakes et al. 1992 and Slavin 1987 for excellent overviews). On the one hand, all these types of curriculum differentiation share the common purpose of tailoring instruction to students with different skill levels. On the other hand, they are different organizational arrangements and should be studied as such. It is therefore problematic when, as noted at the outset, terms are used interchangeably – especially when conclusions from research on one type of curriculum differentiation are generalized to other types.

As one example, Loveless (1998) drew from two meta-analyses that have been conducted, one of which was Slavin’s (1987) study. Loveless (1998) concluded that in elementary school, “[a]bility grouping promotes achievement, and no particular group of children – high, middle, or low ability – misses out on the gain.” Yet, Slavin’s (1987:320) own statement on within-classroom grouping for reading was that “[t]here is not enough research on within-class ability grouping in reading or in the primary grades to permit any conclusions.” Loveless’s (1998) statement fails to make distinctions between research on reading and math, and in the earlier versus later years of elementary school, and therefore overgeneralizes and adds to the confusion.

This point also exemplifies the second barrier to drawing confident conclusions – there is, in fact, a relative lack of research on within-classroom grouping for reading.
instruction. Slavin (1987), quoted above, is not the only scholar to point this out. Ferguson (1998:366), for instance, added 11 years later that “[t]here is no methodologically sound evidence on the effects of within-class ability grouping for reading.” No such evidence has emerged in the literature since Ferguson (1998) reached his conclusion.

A third issue is that the studies that do exist have been limited in important ways, making it even more difficult to draw conclusions. Findings from some of the better studies suggest that, among students whose teachers group, those placed into higher-ranked groups learn more than those placed into lower-ranked groups. In a study of fourth graders in Texas, for example, students in higher skill groups learned more than those in lower groups net of race, social class, prior achievement, and peer and classroom contextual factors (Rowan and Miracle 1983). Gamoran (1986) found evidence that first graders placed into higher skill groups learn more new words than their lower-ranked peers, controlling for age, gender, SES, and a measure of reading aptitude. Another study (Pallas et al. 1994) found positive effects of group rank on first graders’ reading marks net of prior achievement and other controls among a sample of first graders followed through fourth grade. There is thus some limited evidence suggesting that students in higher-ranked groups learn more than their lower-ranked counterparts.

The question is, why? Researchers have explored whether skill grouping matters through variations in peer context, institutional status of the groups, educational expectations, or differences in instruction (Barr and Dreeben 1983; Gamoran 1986; Pallas et al. 1994; Rowan and Miracle 1983). While evidence for these hypotheses has been
mixed, the most consistent finding has been that differences in the quantity and quality of instruction across the group ranks account for much of the group rank effect. When Gamoran (1986), for example, included measures of words taught and phonics taught in his model predicting words learned, the effect of group rank became nil midway through the school year, suggesting that differences in instruction across the groups were the driving force behind the effect of group rank. Barr and Dreeben (1983) also emphasized the importance of instruction, pointing out that teachers create groups for the purpose of providing differential instruction in the first place.

As noted above, however, these studies have been limited in key ways, and their drawbacks necessitate additional and careful consideration of whether skill grouping influences students’ learning. First, the data used in past studies have been limited in terms of sample size and generalizability. Studies of early within-classroom skill grouping have typically relied on small samples, usually drawn from a handful of schools within one or a few districts in a single geographic area. Rowan and Miracle’s (1983) sample, for instance, consisted of 148 students from six schools in one district in Texas. Gamoran’s (1986) study used data on three districts, six schools, and a total of twelve classrooms in the Chicago area; his sample size was approximately 250. Pallas et al. (1994) drew from nineteen schools within a single urban district (Baltimore), with a total sample size of 756. There are representative data on secondary students in the U.S. that have been used to study high school course taking, such as the National Education Longitudinal Study (NELS) and High School and Beyond (HSB), but few studies have
analyzed skill grouping in the early elementary years with nationally representative data – because such data have not been available.

As a result of these data limitations, past studies have been unable to make the kinds of apples-to-apples comparisons needed to demonstrate that students’ learning is really influenced by the skill group into which they are placed. When studies find that high-grouped students learn more and low-grouped students learn less, for example, it is difficult to dismiss the possibility that such patterns represent selection effects rather than causal effects. That is, such patterns of learning might occur regardless of group placement because the kinds of students that are likely to learn more might be selected into higher-ranked groups and the kinds of students that are likely to learn less might be selected into lower-ranked groups. What might account for these potential selection effects?

One source of uncertainty stems from the fact that researchers have often lacked the data to compare the learning of grouped students to similar students in classrooms in which teachers do not group. Skill group placement might indeed increase disparities between low- and high-achieving grouped students during the school year, but this tells us little about the overall effects of skill grouping unless we make comparisons to students whose teachers do not group for instruction. It might be that disparities in learning between low- and high-achieving non-grouped students emerge similarly to those that apparently occur due to the effects of skill grouping. If so, the apparent effects of group placement on learning found in prior studies may actually be spurious (i.e., Gamoran 1986; Pallas et al. 1994).
Although few have had the data necessary to compare grouped and non-grouped students, some studies have found that using skill grouping does not result in higher average achievement among grouped students compared to non-grouped students. Sorensen and Hallinan (1986), for instance, reported no differences in the learning outcomes of grouped and non-grouped students on average, net of initial skills, race, gender, and the size and racial composition of the classroom. This approach suggests no overall benefit of grouping for students’ learning outcomes and is consistent with the spurious argument. However, others have suggested that similar average achievement among grouped and non-grouped students may be masked by greater gains among high-grouped students and fewer gains among low-grouped students that occur simultaneously and contribute to inequality (Gamoran 1992).

To resolve the issue of whether grouping matters, studies should compare the learning of students placed into low, middle, and high skill groups to non-grouped students who are similar in terms of initial skill levels and a variety of other individual and school attributes that may be related to both grouping and learning. Ideally, researchers could compare the learning of students randomly assigned to various grouped and non-grouped contexts, but since this is not possible, the next best comparison is to test whether a student placed into a certain skill group learns more or less than a similar student whose teacher does not group. This is one approach taken in this study. First, I test for significant differences between grouped and non-grouped students’ in-school environments (class size, teacher attributes, etc.) and, when predicting reading gains, control for these differences when they are apparent. Second, I test for effects of group
rank on reading gains by comparing students placed into the lowest, middle, and highest
skill groups in their classroom to similar students (in terms of initial skills, SES, etc.) in
non-grouped classrooms. These comparisons significantly improve upon prior studies
and lessen the likelihood that selection effects are occurring.

However, selection effects may also be due to unmeasured differences in
students’ out-of-school environments. Sociologists have demonstrated that variations in
children’s out-of-school environments rank high among the factors that shape learning in
school (see Alexander 1997). Yet researchers are rarely, if ever, able to adequately
capture the dynamics of children’s lives outside of school when analyzing the
determinants of learning in school. Although we make efforts to control for various
factors, such as family structure, race, and SES, when modeling learning, such measures
undoubtedly fail to capture all the variation between students that shapes learning.
Factors such as parenting styles and parents’ involvement in children’s education (Lareau
1987, 2002), family wealth and its associated benefits (Oliver and Shapiro 1995; Orr
2003), the many attributes of the neighborhoods in which children live (Ainsworth 2002),
and even genetic potential (Guo and Stearns 2002) are examples of the kinds of things
that are difficult to measure yet vary dramatically and influence learning. The
importance of potential selection effects in research on determinants of educational
performance has been a concern since the time of the Coleman Report (Coleman et al.
1966), and it remains so today.

To address this possibility, I compare how much a given student learns during a
school year in which he or she is grouped to how much that same student learns during a
school year in which he or she is not grouped. This cross-sectional time-series approach offers an even more rigorous test because rather than trying to statistically equalize different students along several dimensions by controlling for various indicators, it compares learning within students and thereby makes unmeasured differences between students a non-issue.

Skill Grouping and Inequality in Test-score Distributions

The aforementioned methods are significant improvements over past attempts to discern whether skill group placement influences learning. By comparing the learning of similar students (or the same student) in different groups, it is possible to estimate how much learning can be attributed to group placement.

If skill grouping is indeed salient, it should follow that the test-score distribution of grouped students becomes more unequal over the course of the school year than that of non-grouped students. That is, if high-grouped students experience an acceleration in their learning while low-grouped students experience a deceleration in their learning, relative to non-grouped students, there should be more cases at the high and low ends of grouped students’ test-score distribution. The end result should be more variance (or greater dispersion resulting from more cases lying far from the mean) in the distribution of test scores among grouped students. Figure 1 graphically depicts this possibility and illustrates the predictions of the different theoretical perspectives.

The bars represent hypothetical distributions of test scores for non-grouped and grouped students; these bars grow longer from fall to spring as more variation in learning outcomes emerges over time. On the far left are non-grouped students, where the arrows
have parallel slopes depicting roughly similar learning trajectories for students of all skill levels. In the middle are grouped students, where the arrows again have parallel slopes—but steeper slopes, representing the benefits of being grouped compared to not being grouped as predicted by the equality of opportunity perspective. On the right are grouped students again, where the arrows do not have parallel slopes but instead differ depending on the group into which a student is placed—as predicted by the reproduction of inequality perspective. If group rank matters, high-grouped students should surpass the learning outcomes of similar non-grouped students while low-grouped students learn less, with middle-grouped students somewhere in between. Furthermore, the variation in test scores should grow larger among grouped students, represented by the longer black bar compared to the gray bar, which is the expected variation for non-grouped students.

This possibility has been hinted at—yet rarely tested—in the literature. For instance, Gamoran (1992:13), in a synthesis of research, concluded that “grouping and tracking rarely add to overall achievement in a school, but they often contribute to inequality.” Testing this possibility is statistically simple, yet few studies have made such a comparison—primarily due to the lack of data on non-grouped students, noted above. One exception is the work of Hallinan and Sorensen (1983), who found that the variance in test scores was greater in grouped classrooms than in non-grouped classrooms, but only when the groups were very homogeneous in terms of students’ skills.

To address this issue, I use the standard deviation from the mean reading test score in both the fall and spring of the school year for grouped and non-grouped students.
to determine which category of students exhibits a greater increase in test-score variation during the school year. A greater increase in the standard deviation of grouped students’ test-score distribution relative to that of non-grouped students would be consistent with the notion that skill grouping promotes unequal learning outcomes. My data are not detailed enough to determine the homogeneity (or any other characteristics) of specific skill groups – my sample comes from a few students within each of thousands of classrooms in schools across the U.S. – so it is possible that these analyses will be too aggregate to reveal a distinct pattern in the bivariate.

**Skill Grouping’s Role in Racial and Socioeconomic Gaps in Learning**

Terms such as “the achievement gap” and “the Black-White test score gap” (Jencks and Phillips 1998) have become commonplace both among researchers and in the education policy arena. Poor and minority students consistently perform at lower levels on standardized tests of academic achievement, and education policy – especially at the federal level – has long made attempts to close such gaps (i.e., Head Start, Title I funding, etc.). How might skill grouping play a role in shaping these gaps?

As noted above, poor and minority students tend to be over-represented in lower-ranked groups, and it is possible that students in lower-ranked groups experience decelerated learning compared to similar non-grouped students. Considering these two possibilities together, the concern from the *reproduction of inequality* perspective is that poor and minority students’ disproportionate lower group placement plays an important role in maintaining or even exacerbating achievement gaps between them and their more advantaged peers (Entwisle and Alexander 1993; Gamoran et al. 1995). Existing studies
have tended not to directly analyze the effect of early elementary skill grouping on racial and socioeconomic gaps in learning; thus the notion that skill grouping causes poor and minority students to fall further behind their advantaged counterparts remains much more of a speculation than an empirically demonstrated pattern. As one example, Gamoran et al. (1995:709) stated that “[a]bility grouping divides students on social as well as cognitive characteristics, so by magnifying achievement inequality it contributes to overall achievement inequality among social groups.” Yet, Gamoran et al. (1995), and most researchers studying skill grouping, lacked the data to actually test whether students from disadvantaged backgrounds lose more ground when they are grouped versus not grouped. Such an analysis would significantly improve our understanding of the effects of grouping on disadvantaged students.

In this study, I analyze the impact of skill grouping on racial and socioeconomic gaps in learning in two ways. I first test whether racial and socioeconomic gaps in reading gains grow more unequal over the course of the school year among grouped students than among non-grouped students. I then test whether skill group placement mediates racial and socioeconomic gaps in learning that are evident when grouping is unaccounted for. Reductions in the magnitude of race and SES coefficients once grouping is controlled would suggest that racial and socioeconomic gaps in learning are partially attributable to differences in skill group placement. If racial and socioeconomic gaps are unresponsive to the addition or subtraction of skill grouping to/from the model, then it is difficult to attribute these gaps to the use of skill grouping.
Research Questions and Hypotheses

The findings and limitations of past studies of within-classroom skill grouping lead to several research questions, each of which generates competing hypotheses – one from the equality of opportunity perspective and one from the reproduction of inequality perspective. Below, I outline the specific research questions addressed in this study and identify the predicted answers to those questions from both theoretical perspectives.

(1) Are racially and economically disadvantaged students disproportionately placed into lower skill groups, and, if so, does this pattern come about due to potential discrimination on the part of teachers or rather pre-existing gaps in skills and teacher-reported effort by race and SES? From the equality of opportunity perspective, group placement should be based on students’ skills and effort, not ascribed characteristics such as SES and race. In other words, the pace and level of difficulty of instruction should be appropriately tailored to students based on merit, not ascribed characteristics. From the reproduction of inequality perspective, group placement may be more loosely based on students’ skills and effort because of bias on the part of teachers in assigning students to groups. Ascribed characteristics such as SES and race will be related to group placement even after skills and effort are taken into account.

(2) On the whole, do grouped students learn more than non-grouped students, and does the skill level of the group into which a student is placed matter? From the equality of opportunity perspective, grouped students, on average, should learn more than non-grouped students. This would suggest that skill grouping is a rational and efficient method of managing heterogeneity in students’ skills that helps all students learn more.
This hypothesis also implies that students placed into high, middle, and low reading groups should all learn more than similar students who were not grouped. From the reproduction of inequality perspective, grouped students, on average, will learn less than non-grouped students. This would suggest that skill grouping is not a rational and efficient method of managing heterogeneity in students’ skills, and that all students do not learn more in this context. This hypothesis also predicts that students placed into high skill groups learn more, while students placed into low skill groups learn less, than similar students who are not grouped.

(3) Does inequality in reading skills grow more during the school year among grouped students than among non-grouped students? From the equality of opportunity perspective, there should be less growth in inequality in learning outcomes among grouped students than among non-grouped students. This would be consistent with the view that schools reduce inequality. From the reproduction of inequality perspective, the prediction is more growth in inequality in learning outcomes among grouped students than among non-grouped students. This would be consistent with the view that schools, through skill grouping, exacerbate inequality.

(4) Does skill grouping increase racial and socioeconomic gaps in learning compared to those which are present among non-grouped students? From the equality of opportunity perspective, disparities in learning between students of different social origins should be less pronounced when grouping occurs compared to when grouping does not occur. This is consistent with the notion that schooling allows disadvantaged students to close test-score gaps between themselves and their higher-status peers. From
the reproduction of inequality perspective, disparities in learning between students of different social origins will be more pronounced when grouping occurs compared to when grouping does not occur. This is consistent with the notion that schooling contributes to test-score gaps between students of different social origins and therefore reproduces inequality.
To address the research questions outlined above, I use data from the best source of information on young children and their early childhood experiences currently available: The Early Childhood Longitudinal Study-Kindergarten Cohort (ECLS-K), which was sponsored by the U.S. Department of Education’s National Center for Education Statistics (NCES) and conducted by Westat, the Survey Research Center and the School of Education at the University of Michigan, and Educational Testing Services (ETS). The ECLS-K is a nationally representative sample of 21,260 students who began kindergarten in the fall of 1998 and were followed through the fifth grade (spring 2004).

Teams of trained researchers conducted one-on-one assessments of the children’s skills in reading, math, and general knowledge, and also collected information from the students’ teachers, school administrators, and parents. Data were collected on the full sample in both the fall and spring of kindergarten, on a 30% subsample in the fall of first grade, and on the full sample again in the spring of first, third, and fifth grades. The multi-stage design of the study was fairly complex, as students were sampled within both public and private schools with kindergarten programs, which were sampled within 100
primary sampling units, which were sampled within an initial pool of 1,335 regions (usually counties; for details, see National Center for Education Statistics 2001).

Sample

Most of my analyses focus on the first-grade year. Since I use data from both the fall and spring waves, the sample is limited to the 27% subsample for which both fall and spring information is available. Once these students are selected and repeaters of kindergarten are excluded, my sample consists of 4,953 first graders.

Why focus on first grade and not kindergarten or third grade? Several issues guide my decision. First, entry into first grade is considered to be the beginning of the formal curriculum (see Entwisle and Alexander 1993). Children are undoubtedly exposed to various educational opportunities in kindergarten (and prior to kindergarten), but these experiences vary considerably. For example, 44% of kindergartners in the U.S. attend half-day kindergarten while 56% attend full-day kindergarten (Walston and West 2004). Second, as the ECLS-K data reveal, teachers use textbooks for reading instruction almost universally in first grade; far fewer of the students’ teachers report using textbooks during kindergarten. Third, also evident in the data is the fact that children learn more in first grade than they do in kindergarten (see U.S. Department of Education 2002a). Fourth, teachers report using skill grouping for reading instruction much more frequently in first grade than in kindergarten. Finally, I do not analyze the third-grade data because students’ skills were assessed only once that year – in the spring – meaning that it is not possible to measure school-year learning (an issue I discuss below). For these reasons, I address my research questions using the first-grade data. At various
times, I incorporate data from other waves and report findings that reinforce those from
the first-grade analyses.

**Dependent Variables**

The dependent variables used in this study are measures of reading skills that students
learn over the course of the school year. ECLS-K personnel conducted one-on-one
assessments of students’ skills in five areas of learning that represent a progression of
skills for students at this age: letter recognition, beginning sounds, ending sounds, sight
comprehension of words, and comprehension of words in context. Measuring and
predicting learning over time is a strategy that improves upon cross-sectional estimates of
skills (such as those obtained by states’ proficiency tests) in at least two ways.

First, while a test score from one point in time certainly tells us something about
how much a student knows, it tells us little or nothing about how much a student *learns*.
For an analysis of whether a school process shapes *learning*, such as that being
undertaken here, two test scores are needed – one before and one after exposure to the
independent variable. Second, focusing on learning *during the school year* excludes the
gains or losses that students experience during the summer, when school is not in session.
By measuring and predicting school-year learning, it is more plausible to attribute *gains in skills* to *exposure to schooling*. Along with these benefits, however, come some
methodological concerns with the use of change scores as dependent variables (see, for
instance, Allison 1990; Morgan and Sorensen 1999; Sorensen and Morgan 2000; Willett
1988). In Appendix B, I review several established approaches to measuring and
modeling learning and discuss in more detail the strategies I employ as they relate to
issues specific to the ECLS-K data.

To measure reading gains, I use Item Response Theory (IRT) scale scores
constructed by ECLS-K. Students were first given a routing test that determined which
form (i.e., the difficulty level) of the second-stage reading assessment they would be
given. After both the routing and second-stage tests were scored, all students’ reading
skills were placed on the same metric scale (0 to 92) by using “the pattern of right,
wrong, and omitted responses to the items actually administered in a test and the
difficulty, discriminating ability, and ‘guess-ability’ of each item” (National Center for
Education Statistics 2002:3-2). The IRT scale scores are thus not measures of the
number of correct answers; rather, they “represent estimates of the number of items
students would have answered correctly if they had taken all of the 92 questions in the
first- and second-stage reading forms...” (National Center for Education Statistics
2002:3-4). The IRT method allows lower achievers more opportunities to answer easy
questions incorrectly and high achievers more opportunities to answer difficult questions
correctly – in other words, IRT reduces the likelihood of “floor” and “ceiling effects.”
This approach also “makes possible longitudinal measurement of gain in achievement
over time, even though the tests administered are not identical at each point. The
common items present in the routing test and in overlapping second-stage forms allows
the test scores to be placed on the same scale, even as the two-stage test design adapts to
children’s growth over time” (National Center for Education Statistics 2002:3-3).
*Fall reading skills* is the IRT score for the test taken in the fall of first grade (y1), and *reading gains* is the difference between the fall and spring IRT scores (y2–y1). I also measure students’ *percentage of maximum possible gain*, or how much students learned relative to how much they could have possibly learned (((y2–y1)/(92–y1))*100). Ranges, means, and standard deviations for these and all other measures used in this study are presented in Table 1, and Appendix B provides more information on the two dependent variables.

**Measures of Skill Grouping**

Measures of skill grouping are based on two questions asked of teachers: (1) How many achievement groups in reading do you currently have in this child’s class? (None, One, Two, Three, Four, or Five or more). (2) In which reading group is this child currently placed? (One [highest], Two, Three, Four, Five, Six, Seven, Eight and above, or Not applicable).

To calculate *skill group rank*, I followed the approach of Gamoran (1986) and Pallas et al. (1994) in creating a scale that ranks each student’s group in relation to the number of groups in his or her classroom (thereby accounting for the fact that different teachers use different numbers of groups). The advantage of this approach is that it “indicates the relative status of the group in which the student was placed and allows for comparisons across classrooms with different numbers of groups” (Pallas et al. 1994:32) – an important step since I am using data on students from schools and classrooms across the country. Students in the lowest-ranked group out of four are coded 1.25; the remaining values are 3.75, 6.25, and 8.75 from lowest to highest. Students in the lowest-
ranked group out of three are coded 1.67, the middle group is 5.00, and the highest group is 8.33. When there were only two groups used, the values are 2.50 and 7.50. If teachers reported using “Five or more” groups, it is not possible to know the relative group placement of students in that classroom since it is not clear whether there are exactly five groups or some other number of groups greater than five. Students in these classrooms are therefore initially assigned a missing value on the skill group rank variable.

I also use a series of dichotomous indicators of skill grouping. The first, grouped, indicates whether the student’s teacher reported using skill groups for reading instruction. If the teacher reported a group number for the focal child, a value of 1 was assigned to indicate that the child was grouped. If the teacher answered “not applicable” to the question of which group the student was in, a value of 0 was assigned to indicate that the child was not grouped. The remaining four skill grouping variables indicate whether students were placed into the lowest skill group, the middle skill group, or the highest skill group in their classroom (dichotomous indicators based on the group rank scale described above) or were in a non-grouped classroom (the opposite of the grouped indicator and the reference category when this series of dichotomies is used in models predicting learning).

**Other Measures**

I use several other measures in the analyses, which capture variations in students’ individual, school, teacher, and classroom characteristics (listed in Table 1).
Students’ Background Characteristics

ECLS-K constructed a standardized scale of students’ socioeconomic status based on parents’ level of education, occupational prestige, and family income, as well as a quintile coding of this scale that I recode into a series of dichotomous indicators of SES quintile (for details, see National Center for Education Statistics 2002). I use categorical measures of race indicating whether the student is White, Asian, Black, Hispanic, or another racial/ethnic group.

For family structure, I compare students living with both biological parents to all other family types (more detailed comparisons did not yield significant differences in skill group placement or learning between more specific family types). Female is a dichotomous variable coded 0 if male and 1 if female. ECLS-K measured age in months by dividing by 30 the number of days between the child’s date of birth and the date the child was assessed in the fall. I also measure whether the student has a disability (0 = no, 1 = yes). Approaches to learning is a scale measuring the child’s attentiveness, task persistence, eagerness to learn, learning independence, flexibility, and organization (reported by the teacher). This measure is intended to capture students’ overall effort and attitude toward schooling, an important factor that could influence both group placement and how much a student learns. Externalizing problem behaviors is another scale based on teacher reports; it measures the frequency in which the child argues, fights, gets angry, acts impulsively, and disturbs ongoing activities – again, factors that may be related to both skill group placement and learning to read.
School characteristics

Measured school characteristics include the percentage of economically disadvantaged students (a continuous variable indicating the percentage of students eligible for free lunch), the percentage of racial minority students (1 = 0% to less than 10%, to 5 = 75% or more), total enrollment (1 = 0 to 149, to 5 = 750 or more), and an indicator of private school sector (0 = public, 1 = private).

Teacher characteristics

I measure whether the student’s teacher is non-White (0 = no, 1 = yes), the teacher’s age (in years), level of education (1 = high school or associate’s degree, to 5 = education specialist/professional degree/doctorate), and teaching experience (number of years).

Classroom characteristics

I use several other measures based on teacher reports of the student’s classroom climate: total number of students, percentage of racial minority students, adequacy of textbooks (1 = never adequate, to 4 = always adequate), adequacy of classroom space (1 = never adequate, to 4 = always adequate), and the students’ behavior as a group (1 = misbehave frequently, to 5 = behave exceptionally well).

Handling Missing Data

Researchers have recently begun to pay increasing attention to the various strategies for addressing the problem of missing data (i.e., Allison 2002). There are several ways to handle missing data, each with its own benefits and drawbacks. Here, I briefly review some of these strategies and discuss those that I employ in this study.
One set of methods does not attempt to replace missing values with a feasible alternative to the actual value. *Mean substitution*, for example, involves replacing a missing value with the mean value of that variable. This allows one to retain the original sample size, but is clearly not an accurate way of estimating the missing value. *Listwise deletion* of cases with missing values is another method, which involves excluding any case with a missing value on any variable involved in the analysis. This approach can produce unbiased estimates if values are missing completely at random (i.e., the fact that the value is missing has nothing to do with what the actual value is), but it produces larger standard errors since less information is utilized. *Dummy variable adjustment*, where a dummy indicator of whether a value is missing is included in the model, makes full use of the data but produces biased estimates (Allison 2002).

Another set of methods, however, attempts to make use of available information – i.e., other variables – to predict what a missing value might be. *Imputation*, for example, draws on the relationships between variables in a specified model to predict an alternative to the missing value. Imputation makes full use of the data by restoring the original sample size, and it better estimates missing values relative to mean substitution. However, there is uncertainty as to whether the imputed value is the actual missing value, and this uncertainty results in underestimated standard errors and overestimated test statistics. *Multiple imputation*, on the other hand, is an approach that imputes missing values several different times in order to introduce random variability into the imputation process and calculate standard errors that are not downwardly biased as they would be in the case of a single imputation. Analyses are then performed on each data set separately.
and the results of each analysis are combined to generate the final estimates. Multiple imputation thus allows one to make full use of the data while overcoming the uncertainty associated with relying on a single imputation (Allison 2002).

In the analyses that follow, I use both listwise deletion and multiple imputation. When few variables are involved, and, therefore, the loss of cases due to missing data is not a major concern, I use listwise deletion of cases with missing values. In analyses in which several variables are involved, however, listwise deletion results in a considerable loss of cases. Some models, for instance, include 10 to 20 variables that represent information collected from various sources (i.e., students, teachers, principals, and parents), which increases the chances for a case to be excluded. Therefore, I generate estimates based on five imputations of missing values. Throughout the presentation of results, I note the strategy for handling missing data that is employed for each analysis.

**Accounting for the ECLS-K Sample Design**

Because of the ECLS-K sample design, with two strata (public- and private-sector schools) and students sampled within schools, there is likely to be less variation in the types of children sampled than would have been observed in a simple random sample drawn from the population of all kindergartners in 1998-99. Traditional ordinary least squares (OLS) regression analyses performed on survey data such as these produce under-inflated standard errors and thus increase the likelihood of (incorrectly) rejecting a null hypothesis. To adjust standard errors accordingly when appropriate (i.e., in the regressions predicting skill group rank and reading gains), I use SAS PROC
SURVEYREG, taking into account the two strata and the clustering of students within schools. This procedure uses generalized least squares (GLS) estimation.

Since students are nested within schools in the ECLS-K, hierarchical or multilevel modeling techniques could also be employed. My purpose, however, is to use the broad scope and generalizability of the ECLS-K sample to establish that the effects of skill grouping on learning are not really due to unmeasured differences between students that are associated with selection into groups. I therefore account for the survey design in my analyses and present the results of the GLS models. Testing whether the effects of skill group placement on learning vary across schools or classrooms (i.e., whether there are random effects) is a separate question that could be tested using hierarchical or multilevel models provided that researchers have theoretically-grounded hypotheses and interpretations. I did, however, run supplemental two-level hierarchical models, with the effects of all independent variables fixed across schools. Results from these models are similar to those produced by the GLS models presented here.

Another issue that must be taken into consideration is the timing of the fall and spring skill assessments. Since these assessments occurred over a period of a few months (from September to November in the fall and from April to June in the spring), some students had as few as 146 days between assessments in first grade, while others had as many as 281 (Table 1). To capture the potential impact of this variation on students’ learning, I include a control for number of days between assessments in all models predicting reading gains.
**Analytic Strategies**

The analyses proceed in three major steps, and I present results in three separate chapters. First, I test whether socioeconomically disadvantaged and racial minority students tend to be placed into lower-ranked skill groups than their advantaged counterparts, and, if so, whether these patterns come about due to individual differences in effort, behavior, and reading skills, or rather potential bias or discrimination on the part of teachers (Chapter 4).

Second, I take two approaches to testing whether skill group placement influences students’ reading gains – one focusing on first grade that includes a wide range of control variables in order to statistically equalize different students along several important dimensions, and another that makes within-student comparisons of reading gains across kindergarten and first grade among those who happened to be grouped one school year but not the other (Chapter 5). In the third step, I compare the inequality in the test-score distributions of grouped and non-grouped students and test whether skill grouping plays a role in racial and socioeconomic gaps in reading gains (Chapter 6).
CHAPTER 4

DISPARITIES IN SKILL GROUP PLACEMENT

When teachers create groups of similarly-skilled students for reading instruction, do characteristics other than the students’ skills influence teachers’ decisions? Past research has found a consistent pattern whereby poor and racial minority students tend to be more often placed into lower-ranked groups than their higher-SES and White counterparts. Less clear is whether this pattern comes about due to fewer skills on the part of disadvantaged students or, rather, bias on the part of teachers. Identifying which of these is the primary mechanism driving group placement is directly relevant to the theoretical perspectives on stratification and education outlined earlier. If disadvantaged students are placed lower simply because they have fewer skills, then homogeneous grouping may be a neutral practice – consistent with the equality of opportunity perspective. If disadvantaged students are placed lower on the basis of their ascribed characteristics, however, this may suggest bias or discrimination on the part of teachers – consistent with the reproduction of inequality perspective.

In the analyses that follow, I begin by testing whether poor and minority students are indeed overrepresented in lower-ranked reading groups in the ECLS-K data, as they are in several prior studies. This involves two generalized least squares (GLS)
regressions predicting the continuous skill group rank scale, one with only socioeconomic status included and one with only race included. These initial models establish whether there are socioeconomic and racial disparities in skill group placement at the bivariate level. The expectation, based on past studies, is that lower-SES and minority students will be placed lower on average than higher-SES and White students.

While my focus is on potential socioeconomic and racial disparities, other ascribed characteristics might also influence teachers’ grouping decisions. In a third equation, therefore, I add family structure, gender, age, and whether the student has a disability to the model. Students from two-parent families, girls, older students, and students who are not afflicted with disabilities may be at an advantage when it comes to teachers’ grouping decisions. Moreover, variations in family structure and disability status in particular may account for potential socioeconomic and/or racial disparities evident in the bivariate.

In a fourth equation, I introduce a second set of controls measuring students’ effort, behavior, and skills: approaches to learning, externalizing problem behaviors, and fall reading skills. Students whose teachers believe they put forth more effort may be more often placed into higher-ranked groups, while students who misbehave more frequently may tend to be placed lower. The student’s level of reading skills, of course, is expected to be the primary determinant of skill group placement, with highly-skilled students being placed higher and lower-skilled students being placed lower. If SES and race are no longer predictors of skill group placement once these measures are controlled, it is likely that teachers are not discriminatory on a widespread basis and do indeed place
students into groups in a meritocratic manner. If associations between ascribed characteristics and group placement uncovered in the first two (or three) models persist even net of students’ effort, behavior, and skills, teacher bias becomes a concern.

In all models, I also include measures of the composition of the school’s student body. Because of the likelihood that students of different racial and socioeconomic statuses have different chances for high or low group placement depending on the racial and socioeconomic composition of the school they attend, I control for the school’s percentage of economically disadvantaged and racial minority students in all models. For example, Black-White disparities in group placement cannot occur in all-Black or all-White schools; similarly, we should not expect to find socioeconomic gaps in schools that are socioeconomically homogeneous. Given the segregated nature of schooling in the U.S. (Kozol 1992; Rivkin 1994), capturing these variations is important.

**Results**

The results of these analyses are presented in Table 2. First, Model 1 reveals that students in each of the four lower SES quintiles are placed into lower-ranked groups than students in the highest SES quintile. It is, indeed, a consistent pattern: The more disadvantaged the student’s socioeconomic background, the lower the rank of the reading skill group into which he or she is placed. Moreover, Model 2 shows that Black ($b = - .67; p < .001$) and Hispanic ($b = -.44; p < .05$) students are placed significantly lower than White students (the average group rank of Asian students and those from other racial/ethnic backgrounds does not differ significantly from that of White students). The evidence from Models 1 and 2, then, is consistent with the expectation that poor and
minority students tend to be more often placed into lower-ranked groups than their higher-SES and White counterparts.

Model 3 includes additional student background characteristics and suggests that children from families in which both biological parents are present are placed higher than students from other types of families ($b = .54; p < .001$), girls are placed higher than boys ($b = .61; p < .001$), older students are placed higher than younger students ($b = .04; p < .001$), and students with disabilities are placed lower than other students ($b = -.80; p < .001$). The racial differences lose statistical significance in Model 3, suggesting that family structure, gender, age, and disability status are more salient than race when it comes to disparities in skill group placement. The SES gaps, however, persist and only slightly decline in Model 3.

Turning to Model 4, approaches to learning ($b = 1.37; p < .001$), externalizing problem behaviors ($b = .38; p < .0001$), and fall reading skills ($b = .09; p < .001$) all exhibit large, significant effects on group rank. Students whose teachers believe they are more attentive, persistent, eager, independent, and flexible (approaches to learning) tend to be placed into higher-ranked groups than their peers who are less favorably rated on these dimensions. And, consistent with the underlying purpose of skill grouping, students who begin the school year with more reading skills are placed into higher-ranked groups than students with fewer skills. Once these measures of effort, behavior, and reading skills are included in the model, ascribed characteristics – with a couple of exceptions – tend to be poor predictors of skill group rank.
Unexpectedly, Model 4 suggests that students who act out and misbehave more often (externalizing problem behaviors) are actually placed into higher-ranked groups than their more introverted peers. The opposite might be expected – that teachers place more well-behaved students in higher-ranked groups while poorly behaved students are placed lower so as to not disrupt the lessons of the faster learners. One possible interpretation of this finding is that the students whose teachers rate them as acting out and misbehaving more are also more willing to read out loud and participate in reading lessons, thereby earning them higher group placements. Moreover, the teacher reports are not specific enough to determine whether these students misbehave during skill-grouped reading lessons. It may be that the bad behavior occurs more so during recess or at other times and therefore is not an indicator of behavior in the reading group specifically.

Past studies have suggested that race influences teacher reports of students’ behavior, so controlling for these measures (approaches to learning and externalizing problem behaviors) does not completely rule out the possibility of teacher bias. Studies by Alexander et al. (1987) and Downey and Pribesh (2004), for example, both revealed that Black students’ behavior is evaluated more unfavorably by White teachers than it is by Black teachers. It is possible, then, that teacher bias plays a role in grouping decisions that is not captured by the measures of behavior used here. Still, in supplemental models not shown, SES and race were not significant predictors of group rank even when the measures of behavior were excluded from Model 4. In other words, fall reading skills explain socioeconomic and racial gaps in group rank regardless of differences in
students’ behavior. Teacher bias could be occurring, but it is unlikely to be nearly as important as skills in shaping teachers’ grouping decisions.

A second finding in need of additional interpretation is that students in schools with higher percentages of racial minority students appear to exhibit higher group ranks than students in schools with fewer minority students (Models 2, 3, and 4). This is most likely due to a pattern reported in the next section, that grouped students in ECLS-K are more likely to attend schools with higher percentages of minority students relative to non-grouped students. Therefore, the measure of school racial composition may be artificially associated with higher group rankings because a disproportionate amount of the grouped sample (analyzed in Table 2) attends high-minority schools.

**Theoretical Implications**

What theoretical conclusions can be drawn from the findings reported in Table 2? Although there is some limited evidence that ascribed characteristics shape teachers’ group placement decisions, the findings do not reveal any sort of systematic pattern whereby poor and minority students are placed into lower-ranked groups on the basis of their socioeconomic or racial status. Instead, and consistent with past research, students’ academic skills and effort are the primary determinants of group placement. SES, race, and the other background controls explain only about 8% of the variance in group rank (Model 3); the inclusion of students’ effort, behavior, and initial reading skills increases the amount of explained variance to 42% (Model 4). Consistent with the equality of opportunity perspective, then, socioeconomic and racial inequalities in skill group placement come about not due to discrimination on the part of teachers, but because of
individual students’ varying degrees of effort and behavior, and socioeconomic and racial gaps in reading skills that are already present when first grade begins.

While one conclusion thus might be that placement into skill groups is a meritocratic process, there remains the concern from the *reproduction of inequality* perspective that skill grouping segregates students by SES and race. Although SES and race do not appear to be the *determinants* of differences in group placement, Models 1 and 2 in Table 2 nonetheless show that socioeconomically disadvantaged, Black, and Hispanic students are more often placed into lower-ranked groups than higher-SES and White students. Figures 2 and 3 graphically depict these patterns, which reveal a clear disadvantage when it comes to poor and minority students’ opportunities to learn. As a result of disadvantages in their non-school environments, poor and minority students begin school with fewer skills and, in turn, may more often miss out on opportunities to learn by being placed into lower-ranked skill groups for reading instruction.

Perhaps an even more important question, then, is whether being placed into a particular skill group has an impact on students’ learning during the school year. That is the question I address in the next chapter.
CHAPTER 5

THE IMPACT OF SKILL GROUPING ON READING GAINS

As noted above, the issue of whether students – and which students – benefit from homogeneous skill grouping remains largely unresolved, particularly when it comes to within-classroom grouping for reading instruction during the early elementary years. By comparing the school-year gains in reading skills of grouped and non-grouped students in recent nationally representative data on kindergartners and first graders, the analyses presented below represent significant improvements upon previous attempts to determine whether placement into a particular skill group has an effect on how much students learn. Before turning to these models, however, I begin by comparing the learning environments of grouped and non-grouped students in order to determine whether potential differences make the grouped/non-grouped distinction non-random.

**Differences Between Grouped and Non-Grouped Students’ Learning Environments**

Are students whose teachers use skill grouping systematically different than their peers whose teachers do not group? The answer to this question is important, since comparing the learning outcomes of grouped and non-grouped students is a key contribution of this research. In the ECLS-K sample, roughly 70% of the first-grade students were
homogeneously grouped for reading instruction while 30% were not. I conducted t-tests to determine whether grouped and non-grouped students differ in terms of the various school, classroom, and teacher characteristics discussed previously. These analyses offer insights into what these differences might be, and, therefore, what is important to hold constant in the models designed to isolate the effects of group placement on reading gains.

Table 3 displays the results of the t-tests. Beginning with school characteristics, it is evident that the average percentage of economically disadvantaged students is slightly lower among grouped students than among non-grouped students (28.71 vs. 30.86; $p < .01$). This is contrary to the expectation that disadvantaged schools group more in order to “save a few” of the best and brightest students. The percentage of racial minority students in the school, however, is slightly higher among grouped students (2.84 vs. 2.69; $p < .01$; note that the coding of this variable is not in percentage points). In terms of school size, students whose teachers group do appear to attend larger schools (3.52 vs. 3.22; $p < .001$). Finally, a much lower percentage of grouped students attends private schools than among non-grouped students (.18 vs. .27; $p < .001$).

Surprisingly, there are few differences in grouped and non-grouped students’ classroom characteristics. The number of students in the class, the adequacy of textbooks and classroom space, and the students’ behavior all appear to be unrelated to the use of skill grouping. The exception is the classroom’s percentage of racial minority students, which is notably higher among grouped students than among non-grouped students (36.67 vs. 32.90; $p < .001$). In terms of attributes of the teachers themselves, race and
teaching experience are unrelated to the use of skill grouping, but there is evidence that grouped students tend to have slightly younger (42.90 vs. 43.81; \( p < .01 \)) and more highly educated (3.16 vs. 3.06; \( p < .001 \)) teachers than non-grouped students.

In sum, while there appear to be ways in which grouped and non-grouped students’ learning environments differ, these differences tend to be more random than systematic. There is not a specific pattern that allows for a conclusion as to whether disadvantaged students are grouped more; that appears to be the case when it comes to race (both school and classroom percentage of racial minority students are higher among grouped students than among non-grouped students), but not so when it comes to SES (a smaller percentage of grouped students is economically disadvantaged compared to non-grouped students). Most surprising is that the classroom characteristics one might expect to matter most – since grouping occurs within classrooms – do not appear to be related to grouping (grouped and non-grouped students do not differ in terms of their class sizes, adequacy of textbooks and classroom space, or student behavior). In the next section, I build models predicting reading gains that control for, in addition to various individual background characteristics, those school, classroom, and teacher attributes that were found to be related to the use of skill grouping in Table 3.

**Effects of Skill Group Rank on Reading Gains**

Does the rank of the skill group to which a student is assigned influence how much he or she learns? I address this fundamental question in two ways.
Reading Gains During First Grade

First, I regress both dependent variables (reading gains and percentage of maximum possible gain) on the categorical skill grouping variables described above, controlling for a wide variety of measures capturing variation in students’ home and school environments (Table 4). I estimate learning gaps between both (a) grouped and non-grouped students on average (Models 1 and 3), and (b) students placed into lowest, middle, and highest groups compared to non-grouped students (Models 2 and 4). The general equation is:

\[ y_i = \alpha + \sum_j (b_j X_{ji}) + e_i \]

Reading gains during first grade for student \( i \) are expressed as a function of an intercept, \( j \) independent variables, and an error term. The measures of skill grouping are the key test variables, and the control variables are intended to statistically equalize students as much as possible so as to limit the likelihood that unmeasured differences between students bias the estimates of skill group rank on learning.

Beginning with Models 1 and 3 in Table 4, there appear to be no significant differences in the average reading gains of grouped and non-grouped students. This finding runs counter to the equality of opportunity perspective, which predicts greater reading gains among grouped students as a result of tailoring the pace and difficulty of instruction to the needs of all students. In contrast to this view, I find that skill grouping does not promote a higher overall level of learning than does non-grouped instruction. The question that remains is whether this aggregate comparison masks potentially
divergent learning outcomes among students placed into low- and high-ranked groups, as suggested by the reproduction of inequality perspective.

Models 2 and 4 conduct this test, comparing students in lowest-, middle-, and highest-ranked groups to their counterparts in non-grouped classrooms (the reference category). Beginning with reading gains, it is apparent that lowest-grouped students learn 2.73 fewer points than comparable non-grouped students \((p < .001)\), while highest-grouped students learn 1.25 more points \((p < .01)\). These effects are roughly one third and one seventh of a standard deviation in reading gains, respectively. In Model 4, lowest-grouped students’ percentage of maximum possible gain is 5.00 fewer than that of similar non-grouped students (one third of a standard deviation; \(p < .001)\), while that of highest-grouped students is 3.07 more (one fifth of a standard deviation; \(p < .01)\).

Most of the other results in Table 4 are consistent with what past research would predict. Approaches to learning has a strong, positive effect on reading gains across the board, suggesting that students with greater attentiveness, task persistence, eagerness to learn, learning independence, flexibility, and organization gain more skills than their counterparts who lack these attributes. Older first graders, those with disabilities, and those from lower-SES backgrounds tend to gain fewer points than their otherwise similar peers. In terms of the classroom and school characteristics, students of older teachers appear to gain more than their counterparts with younger teachers, while students in schools with high concentrations of economically disadvantaged and racial minority students gain fewer points than students in higher-SES and White schools. Students with more time between the fall and spring reading assessments gain more points than those
for whom less time elapsed, indicating the importance of sheer opportunity to learn.

Finally, students who begin the school year with more reading skills appear to gain fewer points during first grade than their peers who start off with fewer skills (in Models 1 and 2). The question of why this is the case is addressed in Appendix B; the important point to note here is that the results for percentage of maximum possible gain (Models 3 and 4) are nearly identical to those for reading gains (Models 1 and 2), suggesting that the findings in Models 1 and 2 are robust.

To further illustrate the magnitude of the skill grouping effects, consider the unstandardized coefficient for lowest-group placement relative to that for being in the lowest SES quintile. In Model 2, students in the lowest SES quintile gain 1.87 fewer points than students in the highest SES quintile, but students in the lowest skill group gain 2.73 fewer points than similar non-grouped students. In Model 4, being in the lowest skill group versus not being grouped is again more of a disadvantage than is being in the lowest SES quintile versus the highest SES quintile. Similarly, the highest-group advantage is more pronounced than the advantage of attending a private school in both Models 2 and 4. The effects of group placement – even net of all the controls in the equations – are therefore considerable.

Taken together, the models presented in Table 4 represent important improvements upon previous efforts to discern whether skill group placement shapes learning to read at young ages. Relative to past research, these models are better able to compare the learning outcomes of similar students placed into different instructional contexts for reading lessons. Instead of comparing high-, middle-, and low-grouped
students to one another, my models compare these students to their counterparts in non-grouped classrooms – an approach that reduces the likelihood that selection effects explain why group placement matters. In addition to the skill grouping measures, I control for 17 other factors that may also be related to both skill grouping and reading gains, including students’ initial skills.

Despite this effort, it is still possible that unmeasured differences between students placed into different groups account for the apparent effects of skill group placement on reading gains. Even though I controlled for initial skills and a host of other factors in an attempt to statistically equalize students as much as possible, such an attempt is never perfect because there are, inevitably, factors that cannot be measured. For example, it may be that high-grouped students enjoy reading and are more motivated to learn than low-grouped students, and that is why they learn more. While controlling for things like initial skills and approaches to learning certainly captures some of the variation between students, the extent to which students truly enjoy schooling and are motivated to learn is extremely difficult to measure and may be biasing the estimates of the effects of group placement on reading gains. The next set of analyses addresses this concern by making within-student comparisons.

Cross-Sectional Time-Series Analysis Across Kindergarten and First Grade

The second approach removes between-student variation from the equation by using longitudinal (kindergarten and first grade) data to estimate the effect of being placed into a certain skill group one year compared to being in a non-grouped classroom the other year. To make this comparison, I first determined which students in the fall first-grade
subsample happened to be grouped one year but not the other year. There were 1,909 such students; combining data from both years yielded a pooled, cross-sectional time-series data set with 3,818 student-years. This data set was created, and analyses were performed, based on known information about teachers’ grouping practices; in other words, cases with missing data on any measure used were deleted. The result was a smaller, and probably less representative, sample of students than that used in the previous analyses. The smaller sample size may be a drawback in terms of generalizability, but combined with the rigor of the analyses (discussed below), the result is perhaps the most stringent test for the effects of skill grouping to date.

The cross-sectional time-series analyses, with fixed effects, compare the reading gains of a given student the year he or she was grouped to the year he or she was not grouped:

\[(y_t - \bar{y}_i) = \sum_j b_j (x_{j,t} - \bar{x}_{j,t}) + \eta_t\]

The within-student change in reading gains from one year to the other is expressed as a function of \(j\) within-student (time-varying) characteristics of student \(i\) at time \(t\) and their related error term. The key is that time-invariant factors and their related error term are not part of the equation, meaning that unmeasured differences between students (i.e., family wealth, neighborhood context, genetic potential, motivation, etc.) do not bias the estimates as they do in most models predicting learning (such as those presented in Table 4 above). I exclude 138 students (276 child-years) who changed schools between kindergarten and first grade, thereby making school characteristics essentially time-
invariant across the two years (since all students in the analyses went to the same school both years). Furthermore, I control for an indicator of the kindergarten year to capture additional time-variant factors, i.e., changes in school and classroom characteristics and different rates of learning in kindergarten and first grade. I also include fall reading skills in the model when appropriate (in Models 1 and 2), as well as the number of days between assessments, as time-variant independent variables.

Most importantly, these within-student fixed-effects analyses significantly reduce the likelihood that selection into skill groups explains the apparent effects of skill group rank on learning. In addition to holding constant classroom and school attributes as noted above, whatever characteristics of the children and their out-of-school contexts that are typically not captured by controlling for a variety of socioeconomic and family background measures are controlled in these analyses because they compare the same student in different learning contexts (grouped one year but not the other). I estimate models parallel to those presented in Table 4, first making the grouped/non-grouped comparison and then the lowest, middle, and highest group versus non-grouped comparison (see Table 5).

The results of the cross-sectional time-series models, generated in Stata and presented in Table 5, add considerable merit to the position that skill grouping promotes unequal learning outcomes. First, Models 1 and 3 suggest that being placed into a skill group – regardless of which group – one year does not influence students’ reading gains relative to being in a non-grouped classroom the other year. These results are consistent with those in Table 4 and further undermine the equality of opportunity perspective – a
given student does not appear to gain more reading skills when grouped compared to when not grouped.

Models 2 and 4 are also consistent with the earlier findings, as students learn less the year they were placed into the lowest group, and more the year they were placed into the highest group, compared to the year they were in a non-grouped classroom. The coefficients drop in magnitude and significance levels relative to those in Table 4, but this should not come as a surprise given the smaller N and the rigor of the analysis. The salience of initial skills is again apparent in these models, as is the number of days between the fall and spring assessments. Furthermore, the control for time-variant factors (the indicator of kindergarten vs. first grade) suggests that students gain far fewer reading skills in kindergarten than they do in first grade. The models explain nearly 60% of the within-student variation in reading gains, which is a notably better model fit than if OLS regression were run on the data set without regard to the cross-sectional time-series structure of the analyses (as indicated by the much lower overall R-squares).

Relative Merits of the Two Modeling Strategies

The between-student/GLS and the within-student/cross-sectional time-series models each have their own benefits and drawbacks, but both lead to the same conclusion. The benefit of the between-student models is that they make full use of the sample and produce results that can be generalized to first graders in the U.S. However, the between-student models likely do not capture all of the differences among students, making it difficult to confidently rule out selection effects as at least contributing to the estimated effects of skill group placement on reading gains. The benefit of the within-student
models is that they do account for these potentially unmeasured differences between students by comparing the same student to him- or herself. This significantly reduces the likelihood that selection effects bias the estimates of skill group placement’s effect on learning. The drawback of the within-student models is that they are compromised somewhat by the loss of sample size and generalizability that occurs when I limit the sample to only those students who happened to be grouped one year but not the other.

Importantly, both modeling strategies lead to the same conclusion: Being placed into skill groups results in accelerated learning for high-grouped students and decelerated learning for low-grouped students compared to being in a classroom where the teacher does not group. Both modeling strategies suggest no overall benefit of grouping, but a distinct pattern of divergent learning trajectories for high- and low-grouped students.

**Theoretical Implications**

Taken together, the findings presented in Tables 4 and 5 offer clear support for the reproduction of inequality perspective. Consistent with the prediction illustrated on the far right-hand side of Figure 1, the results reveal unequal, divergent slopes for high- and low-grouped students, not parallel steeper slopes as predicted by the equality of opportunity perspective.

In order to conclude that skill grouping benefits all students, the evidence would have to show that low- and middle-grouped students also learn more than similar non-grouped students. If that were the case, then using skill grouping would be advantageous relative to full-classroom instruction in that it would promote learning in a way consistent with the goal of equal opportunity. In reality, however, the evidence suggests that the
learning trajectories of middle-grouped students are no different than those of similar non-grouped students, while low-grouped students actually gain fewer skills than similar non-grouped students. Skill grouping, then, contributes to inequality by boosting the learning of highly-skilled students but repressing the learning of low-skilled students. I return to these issues in the discussion.
CHAPTER 6

INEQUALITY, SOCIOECONOMIC STATUS, AND RACE

Thus far, the evidence suggests that (a) socioeconomically disadvantaged, Black, and Hispanic students are more often placed into lower-ranked reading groups than their advantaged peers, (b) there are no overall differences in the school-year gains in reading skills of grouped and non-grouped students as a whole, and (c) lowest-grouped students learn less while highest-grouped students learn more than similar non-grouped students.

These findings lead to two other hypotheses. First, since there are divergent slopes for high- and low-grouped students, the amount of variance, or inequality, in the test-score distribution of grouped students should be greater than that of non-grouped students. This would be consistent with the view that skill grouping reproduces (or even exacerbates) inequality in cognitive skills. Second, socioeconomic and racial gaps in learning should be greater among grouped students than among non-grouped students, due to disadvantaged students’ disproportionate placement into lower-ranked groups where less teaching and learning occur. In this chapter, I explore both of these possibilities.
Inequality in the Test-Score Distributions of Grouped and Non-Grouped Students

The findings presented in Chapter 5 show that there is no overall difference in the average learning of grouped and non-grouped students, but masked behind this aggregate comparison is a pattern of divergent learning trajectories whereby high-grouped students learn more and low-grouped students learn less. This implies that the distribution of learning outcomes is more unequal among grouped students than among non-grouped students (as depicted in Figure 1).

In non-grouped classrooms, the tendency should be for students’ skills to become more similar over time, as teachers cover material at the same pace for all students. Accordingly, the variance in the distribution of students’ skills should decrease over the course of the school year. In contrast, the distribution of grouped students’ skills should become more unequal over the course of the school year because teachers cover more material with high-grouped students and less material with middle- and low-grouped students. As noted in Chapter 2, the lack of data on both grouped and non-grouped students has prevented researchers from making this comparison (Hallinan and Sorensen [1983] being an exception).

In order to test these predictions, I conducted analyses that compare the means and standard deviations of the fall and spring reading IRT scores for both grouped and non-grouped students. Figure 4 displays the results. First, looking at the numbers in the bottom-left portion of the figure, the average reading skills of grouped and non-grouped students are similar in the fall, while the standard deviation is larger for grouped students than for non-grouped students ($s = 12.99$ vs. $12.29; p < .05$). This suggests that grouped
students begin the school year with greater heterogeneity in skills compared to non-grouped students, and hints at the possibility that greater heterogeneity in students’ skills is one reason why teachers choose to use grouping. Although the data do not allow me to test this directly, the pattern reported here is consistent with this view and with the underlying theory of homogeneous grouping.

Looking next at the bottom-right of the figure, it is evident that grouped and non-grouped students’ skills are also similar in the spring of the school year, which is not surprising given the results of the analyses comparing grouped and non-grouped students as a whole in Chapter 5. Unexpectedly, however, the standard deviations are not significantly different in the spring. In fact, the standard deviation of grouped students increases less (13.63 – 12.99 = .64) than does the standard deviation of non-grouped students (13.38 – 12.29 = 1.09), the opposite of the prediction. (Supplemental analyses also do not find greater inequality among grouped students when it comes to percentage of maximum possible gain.)

The top portion of the Figure illustrates this pattern graphically. The bars represent the mean plus and minus one standard deviation in reading IRT points. The prediction, as illustrated earlier in Figure 1, was that the bar representing variation in grouped students’ skills would grow longer from fall to spring. Instead, it is apparent that there is very little difference in the extent of inequality in the test-score distributions of grouped and non-grouped students, and that there is not more growth in inequality among grouped students – if anything, inequality grows less among grouped students.
How can these findings be reconciled with those suggesting that high-grouped students learn more, and low-grouped students learn less, than similar non-grouped students? One view is that if skill grouping really matters, there should be more observable inequality among grouped students than among non-grouped students. However, there are several possible explanations for why Figure 4 does not tell the expected story.

The first is that students’ opportunities to move up the IRT scale from fall to spring vary depending on their initial skills, and this phenomenon could have an impact on analyses of variance such as those presented in Figure 4. In Appendix B, I document concerns with a ceiling effect in the reading IRT scale that are relevant here. Many students scored near the top of the scale in the fall of first grade, and those students gained fewer points than students who initially scored lower. Since high-grouped students tend to begin the school year with more skills than middle- and low-grouped students, high-grouped students have less room to move up the IRT scale than middle- and low-grouped students. This creates a pattern in which the variance in high-grouped students’ measured skills is not allowed to increase as much as it might in the absence of a test-score ceiling. I present evidence of this possibility in Figure 5.

I calculated the increase in the standard deviation, from fall to spring, of non-grouped students as a whole, grouped students as a whole, and lowest-, middle-, and highest-grouped students separately. Hidden behind the aforementioned .64-point fall-to-spring increase in grouped students’ standard deviation is a noteworthy pattern: Lowest-grouped students’ standard deviation increases 3.50 points from fall to spring, middle-
grouped students’ standard deviation increases 1.43 points from fall to spring, and
highest-grouped students’ standard deviation decreases by 2.46 points from fall to spring.
In other words, students placed into high-ranked reading groups exhibit less variation in
the spring than they did in the fall. While it is possible that some instructional
phenomenon drives this pattern, the most likely explanation, consistent with the evidence
presented in Appendix B, is that the range of high-grouped students’ skills is not
completely captured by the reading IRT scale. The same is likely to be true for high-
scoring non-grouped students, of course, meaning that the increase in non-grouped
students’ standard deviation is probably also truncated by the ceiling effect. It is
impossible to know whether the truncation in the measurement of students’ skills is
different for grouped and non-grouped students, of course, so it may be that the
seemingly simple comparison presented in Figure 4 does not lead to a reliable
conclusion.

Does this contradict the findings of the multivariate models in Chapter 5, which
suggest that the group into which a student is placed influences learning? It does not,
because the comparisons made here and in Chapter 5 are very different. The multivariate
models estimate the effect of one independent variable (i.e., skill group placement) net of
all other variables included in the equation. Importantly, students’ initial skills are part
of those equations – as a control in the models predicting gains and as part of the
computation of the percentage of maximum possible gain variable. The estimated effects
of skill group placement therefore reflect the expected gains net of students’ initial
scores. This approach neutralizes the impact of a ceiling effect since the effects of the
independent variables on gains are estimated as gains beyond the average initial score. For example, Model 2 in Table 4 estimates that a lowest-grouped student who puts forth the average amount of effort, exhibits the average amount of misbehavior, is the average age, etc., *and who begins the year with the average level of skills*, gains 2.73 fewer points than the same student in a non-grouped classroom. Since the bivariate test presented in Figure 4 is a comparison of two distributions, by its very nature it does not adjust for differences along any other dimensions. In this case, given the potential ceiling effect in the reading scale, the multivariate models are superior to the bivariate comparisons.

Another drawback of the analysis presented in Figure 4 is that the data are not detailed enough to allow for comparisons within classrooms, where differential variance between grouped and non-grouped students would be most expected. A large, national sample like ECLS-K offers the advantages of statistical power and generalizability, but one tradeoff is that the sampled students are scattered across many classrooms. Since data on all students within a given classroom are not available, it is not possible to conduct within-classroom analyses that compare, for instance, the variance in test scores in a grouped classroom to that in a non-grouped classroom. This is what Hallinan and Sorensen (1983) were able to do. Their study found that when the reading groups were very homogeneous (which I cannot measure), grouped classrooms did indeed exhibit greater variance in learning outcomes than non-grouped classrooms. The authors suggested that when students in a given reading group are more similar, instruction is more accurately tailored to the students. In this scenario, the effects of group rank are
more pronounced and there is a greater tendency toward unequal teaching and learning – and, therefore, greater variance among students in grouped classrooms.

All in all, past research suggests that the analysis presented in Figure 4 asks the right kind of question about inequality, but limitations in the data necessitate caution in drawing conclusions based on the results. The test-score scale may not be accurately measuring the higher-scoring students’ skills, and the data are not detailed enough to allow for classroom- or group-level comparisons that might replicate Hallinan and Sorensen’s (1983) findings.

**Socioeconomic and Racial Disparities in Learning: Does Grouping Matter?**

As noted in Chapter 2, two common criticisms of homogeneous skill grouping are that (a) it segregates students by SES and race, and (b) it contributes to socioeconomic and racial inequality in learning. The findings presented in Chapter 4 are consistent with the first criticism – low-SES, Black, and Hispanic students are more often placed into lower-ranked reading groups than their advantaged peers because they begin the school year with fewer skills. The second criticism is the topic of this section: When it comes to socioeconomic and racial disparities in learning, is skill grouping partially to blame?

It is evident in Chapter 4 (and prior studies) that when teachers group, lower-SES and minority students are more likely to be sorted into lower-ranked reading groups for instruction, while higher-SES and White students are more likely to be placed into higher-ranked groups. In non-grouped classrooms, however, students from different socioeconomic and racial backgrounds are not unevenly sorted for reading lessons and are presumably exposed to the same pace/difficulty of instruction. Given the evidence
that high-group placement promotes gains while low-group placement promotes losses – relative to being in a non-grouped classroom – the expectation is that gaps in learning between advantaged and disadvantaged students are larger when teachers group compared to when they do not group. The use of grouping should propel advantaged students toward greater gains while holding back disadvantaged students; at the same time, the tendency should be toward the middle in non-grouped classrooms.

The earlier findings suggest that such patterns might only be evident in the bivariate – that is, when students’ initial skills and other factors are not taken into account. Recall that in Chapter 4, the socioeconomic and racial disparities in skill group rank were only apparent in the bivariate and were explained by differences in students’ fall reading skills, teacher-reported effort, and teacher-reported behavior (Table 2). In other words, students of different socioeconomic and racial backgrounds, but of similar skills, effort, and behavior, are not likely to be unequally placed into skill groups. Moreover, the main models predicting reading gains (Table 4) revealed no significant racial gaps in learning net of the other measures in the equation; socioeconomic gaps were only unveiled when comparing the lowest- and highest-SES students. These patterns suggest that the impact of skill grouping on socioeconomic and racial disparities in learning may prove to be minimal when initial skills and other factors are taken into account. Therefore, I begin with a bivariate analysis.

I first conduct t-tests comparing the average fall and spring reading skills of low-SES students relative to high-SES students, and Black and Hispanic students relative to White students, among those in both grouped and non-grouped classrooms. These
analyses reveal significant gaps in skills in all instances – low-SES students score significantly lower than high-SES students while Black and Hispanic students score significantly lower than White students. These gaps are evident among both grouped and non-grouped students, and in both contexts, the gaps grow larger over the course of the school year. The question is, do the gaps increase more among grouped students than among non-grouped students?

Table 6 displays the growth of socioeconomic and racial disparities in reading gains from fall to spring among grouped and non-grouped students. In grouped classrooms, the reading skill gap between lowest-SES and highest-SES students increases from 15.38 to 17.12, or 1.74 points, over the course of the school year. In non-grouped classrooms, the gap increases from 13.37 to 15.07, or 1.70 points. Contrary to the expectation, then, socioeconomic disparities in learning do not appear to grow much larger among grouped students relative to non-grouped students. The results show similar patterns when it comes to racial gaps. The Black-White gap grows 2.02 points from fall to spring among grouped students, and 2.71 points among non-grouped students – the opposite of the expectation. The Hispanic-White gap also appears to grow less among grouped students (.29) than among non-grouped students (1.86). In pooled GLS regression models not shown, I tested for interactions in order to determine whether the socioeconomic and racial gaps in gains are significantly different for grouped and non-grouped students (i.e., lowest-SES*grouped; Black*grouped; Hispanic*grouped; all in separate bivariate equations). These models find that none of the socioeconomic and racial gaps in reading gains between grouped and non-grouped students are statistically
significant. I also conducted parallel analyses of gaps in students’ percentage of maximum possible gain; these analyses do not reveal larger socioeconomic and racial gaps among grouped students.

Another way to explore the relationship between skill grouping and socioeconomic/racial disparities in reading gains is to test whether estimates of socioeconomic and racial gaps in learning are mediated by skill grouping. I begin with a model that estimates socioeconomic and racial gaps, controlling only for initial skills and the number of days between the two assessments. I then add the measures of skill group placement. Reductions in SES and race coefficients when skill grouping is added to the equation would suggest that some of the socioeconomic and racial disparities in reading gains is attributable to the use of skill grouping. If the SES and race coefficients are unresponsive to this change in the equation, then it is unlikely that skill grouping is driving socioeconomic and racial disparities in reading gains.

Table 7 displays the results of GLS models testing whether skill grouping mediates gaps between low- and high-SES students, predicting both gains (Models 1 and 2) and percentage of maximum possible gain (Models 3 and 4). The inclusion of the skill grouping measures in Model 2 does not reduce the SES coefficients in Model 1. In fact, the gaps appear slightly larger when skill grouping is added to the equation. Looking at the results for percentage of maximum possible gain, however, the SES coefficients do decline from Model 3 to Model 4. The gap between the lowest- and highest-SES students, for example, declines from -11.57 to -9.98 when grouping is included in the equation, suggesting that skill group placement explains 14% of the gap.
Looking at Table 8, a similar pattern is evident for Black and Hispanic students’ gains relative to those of White students (recall that only Black and Hispanic students are placed significantly lower than White students). The models predicting gains do not suggest that skill grouping mediates the Black- and Hispanic-White gaps (Models 1 and 2), but the models predicting percentage of maximum possible gain (Models 3 and 4) hint at the possibility. The Black-White gap, for instance, declines from -7.27 to -6.82, or 6%.

It is difficult to know why the two dependent variables produce different patterns of results in this case. In the main models predicting learning (Tables 4 and 5), results are consistent regardless of which dependent variable is used. In Tables 7 and 8, the results suggest that skill grouping partially explains socioeconomic and racial gaps in percentage of maximum possible gain, but not the standard measure of gains. Given that the bivariate analyses do not suggest a negative role of skill grouping (Table 6), it is difficult to conclude from Tables 7 and 8 that a portion of socioeconomic and racial gaps in learning can be attributed to the use of skill grouping.

Overall, then, the evidence does not implicate skill grouping as a major culprit in the production of socioeconomic and racial disparities in reading gains. The use of grouping does not appear to mediate such disparities; even in the bivariate, there are not larger gaps in reading gains or percentage of maximum possible gain among grouped students relative to non-grouped students. Supplemental analyses (not shown) extending these tests to the multivariate – controlling for initial skills, effort, school socioeconomic
and racial composition, etc. – also failed to produce evidence that grouping is especially
damaging to poor and minority students’ reading gains.

What might be some possible explanations for these patterns? One is that the
effects of skill grouping on learning found in Chapter 5 are spurious. If group placement
truly promotes divergent learning outcomes for low- and high-grouped students, poor and
minority students should lose more ground to their advantaged peers when they are
grouped. Since that does not appear to be the case, it might be that low-group placement
does not hinder learning and that other characteristics of students and their learning
environments are to blame. This explanation is unlikely, though, since the multivariate
models predicting reading gains are the most rigorous to date in terms of isolating the
impact of group placement. These models (Table 4) suggest that grouping is salient, and
even the within-student analyses suggest divergent effects of high- and low-grouped
instruction (Table 5).

A more likely explanation is that socioeconomic and racial disparities in skill
group placement are not severe enough to produce larger gaps in learning than those
occurring in non-grouped classrooms. Recall that in Figure 3, the estimated average skill
group rank of White students is 5.23 on a scale ranging from 1.25 to 8.75. The average
group rank of Black students is 4.56 ($p < .001$) and the average group rank of Hispanic
students is 4.79 ($p < .05$). While these patterns do suggest lower group placement for
Black and Hispanic students relative to White students, and these differences are indeed
statistically significant, they may not be disparate enough to produce significantly fewer
gains in reading skills among grouped minority students compared to non-grouped
minority students. Skill grouping appears to promote divergent learning outcomes regardless of SES and race, and increasing socioeconomic and racial gaps in learning to read may be driven by something other than the use of skill grouping.

It is also possible that the data are not detailed enough to capture instances in which poor and minority students are negatively affected by skill grouping. The type of homogeneous grouping studied here occurs within classrooms, but the data do not include information on each student in every classroom. Such data would allow for analyses that compare socioeconomic and racial gaps in learning within grouped classrooms to those emerging within non-grouped classrooms. Unfortunately, as noted above, students in the ECLS-K sample are scattered across many classrooms. This is especially true of the first-grade wave, since the sample was drawn in the kindergarten year and students then spread out into many different classrooms in first grade. This makes it impossible to analyze entire classrooms of students. The measures of skill group rank used in the analyses do capture the fact that different teachers use different numbers of groups, and that students’ relative group placements vary as the number of groups varies. This allows for excellent analyses of the effects of group rank on reading gains across all students in the sample (i.e., Chapter 5), but the lack of information on all students within classrooms remains a limitation for analyses such as those presented in this chapter.

Theoretical Implications

What conclusions can be drawn, theoretically, from the findings presented in this chapter? In Chapters 4 and 5, the evidence tended to support the reproduction of
inequality perspective. Poor and minority students tend to be placed into lower-ranked reading groups than their high-SES and White counterparts, and placement into lower groups is detrimental to learning compared to being in a non-grouped classroom. In the analyses of inequality among grouped students relative to non-grouped students, the evidence is far less supportive of the view that skill grouping contributes to inequality.

The reproduction of inequality perspective predicts greater inequality in the test-score distribution of grouped students, yet the analyses in this chapter do not reveal such a pattern. Rather, it appears that the extent of inequality is similar among students in both grouped and non-grouped classrooms. Methodological limitations underlie this analysis, though, so it would be premature to conclude that the evidence supports the equality of opportunity perspective.

A similar story emerges when it comes to socioeconomic and racial gaps in learning among grouped and non-grouped students. The reproduction of inequality perspective suggests that disadvantaged students’ lower group placements lead them to lose more ground to their advantaged peers when they are grouped compared to not grouped. Again, the analyses in this chapter do not support this view, but limitations of the data make it difficult to draw confident conclusions.
CHAPTER 7

DISCUSSION AND CONCLUSION

The goal of curriculum differentiation is to tailor the pace and difficulty of instruction to students based on their level of preparedness. Most of what we know about curriculum differentiation comes from studies of students tracked into different courses and/or programs of study at the secondary level, and some academics have long been critical of this practice because it is thought to contribute to inequality (i.e., Oakes 1985). When it comes to within-classroom grouping for reading instruction in the early elementary years, a lack of high-quality studies has prevented researchers from drawing conclusions as to its effects on students’ learning.

This should not come as a surprise, as this type of curriculum differentiation is, arguably, the most difficult type to study. It is implemented at the discretion of individual teachers and therefore varies from one classroom to another. There are no labels such as “college preparatory” or “vocational” to characterize instruction, as with high school tracking. Without actually being in the classroom to observe what is occurring, it is difficult to obtain good data on elementary grouping practices. For these reasons, the relative lack of knowledge about its effects on students is understandable. To reiterate this point, one researcher concluded – as recently as 1998 – that “[t]here is
no methodologically sound evidence on the effects of within-class ability grouping for reading” (Ferguson 1998:366). ECLS-K therefore offers researchers an opportunity to gain a better understanding of group placement and its effects on learning. The current study is an important contribution to the literature, yet it also reveals limitations of studying skill grouping with ECLS-K.

Most notably, my analyses reduce the likelihood that selection effects bias the estimates of the effects of skill group placement on reading gains. Due to a combination of data limitations and the fact that skill grouping is so widespread, most prior studies have focused only on students in grouped classrooms. A handful of studies, for instance, have found that high-grouped students learn more than middle- and low-grouped students. Such analyses actually tell us little, though, because high-grouped students probably would have learned more than their middle- and low-grouped counterparts even if the teacher had not used skill grouping. Controlling for differences between students placed into different groups is not much help, as Slavin’s (1990:505) analogy illustrates: “Are the San Francisco Forty-Niners better than the Palo Alto High School football team, controlling for height, weight, speed, and age? Such questions fall into the realm of the unknowable.”

The literature is in need of analyses that compare the learning of grouped students to that of similar non-grouped students, and my analyses make such a comparison. The findings presented in Chapter 5 suggest that high-group instruction is an advantage, and low-group instruction is a disadvantage, relative to learning in a classroom where the teacher does not group. These patterns persist in both between- and within-student
models, and together, this set of findings represents the most convincing evidence to date that group placement affects learning. How much does skill grouping affect reading gains? If there is one predictor of educational achievement that sociologists have identified, it is that being poor is a disadvantage relative to coming from a higher-SES background. Yet my models (Table 4) suggest that the disadvantage of being in the lowest skill group (versus not being grouped) is greater than that of being in the lowest SES quintile (versus the highest SES quintile). The fact that group placement continues to matter in the cross-sectional time-series models with fixed effects speaks further to the salience of skill grouping (Table 5).

These findings support the reproduction of inequality theoretical perspective, which views education as an institution that reproduces advantages and disadvantages generated outside of school. When teachers use skill grouping, it contributes to unequal learning outcomes: Students with initially fewer skills are placed into lower-ranked reading groups, and those students learn less than they would have in a classroom where the teacher does not group. At the same time, students who begin school with more skills are placed into faster-learning groups, and those students learn more than they would have in a non-grouped classroom. Indeed, there are few educational practices as relevant to the reproduction of inequality as homogeneous grouping, which treats students differently based on their unequal skills and consequently exacerbates those inequalities.

The use of skill grouping warrants more attention from policy makers seeking to reduce inequality in academic achievement. Legislative efforts to raise test scores and close learning gaps have centered around equitable school funding, desegregation,
accountability, and school choice (i.e., charter schools and vouchers), but there is rarely any discussion of skill grouping in the policy arena. Yet, skill grouping is a practice that could be manipulated by policy makers, principals, and teachers interested in reducing learning gaps. The No Child Left Behind Act, for example, emphasizes greater equality in learning outcomes and calls for all students to reach proficiency levels in the next decade. The policy aims to ensure that students are taught by qualified teachers, that parents have school choice, and that students’ progress is monitored with annual proficiency tests, but it does not get at fundamental questions about practices that occur within schools and contribute to inequality in the first place.

Two findings in this study are directly relevant to these issues: (1) Skill grouping does not appear to raise overall levels of achievement compared to non-grouped instruction, but (2) it does appear to promote unequal learning outcomes in the form of greater gains among high-grouped students and fewer gains among low-grouped students. If teachers were to abandon the use of skill grouping, then, there would be about the same overall level of achievement among students, and less inequality in learning outcomes. If skill grouping is not abandoned, the challenge for those seeking to raise test scores and reduce inequality is to find ways to improve teaching and learning in the middle- and low-skill groups so that grouping truly does benefit all students and not just some students (see Gamoran 1992; Hallinan 2003; and for a parallel debate over the merits of curriculum differentiation at the secondary level, see Hallinan 1994a, 1994b and Oakes 1994a, 1994b).
While the goal of the No Child Left Behind Act is apparent in its title, evidence is surfacing that reveals extra help for middle-skilled students, but not lower-skilled students. A recent case study of an elementary school in Texas found that the district’s response to No Child Left Behind was to implement a “data-driven” system whereby students are classified into one of three categories based on practice proficiency test scores: “passers,” “bubble kids,” and “foundation” or “remedial” kids. The passers will likely pass the proficiency test, the foundation or remedial kids will likely not pass, and the bubble kids are close to passing. The bubble kids are given extra attention, such as one-on-one or small group instruction in class and extra instruction outside of class – including after school and/or over the summer. The goal is to use this system of “educational triage” to move bubble kids up to the proficiency level, thereby increasing teachers’ and schools’ overall pass rates (Booher-Jennings 2005). Similar practices have been reported in California (Rubin 2004) and informal discussions with educators suggest that this approach is common and will continue to become more widespread over time.

The assumptions that underlie these practices reveal a great deal about what educators believe shapes learning. The goal of skill grouping is to tailor the pace and difficulty of instruction to students as deemed appropriate for their skill levels; the assumption is that students will not learn effectively if instruction is beyond their “ability.” Educational triage, however, is inconsistent with this line of thinking, instead assuming that “bubble kids” can become “passers” if they are given extra instruction. In this case, the assumption is that students can learn beyond their “ability” if we teach them
more. What is to say that children placed into low-ranked skill groups could not learn more if they were taught more? My findings suggest that their peers in non-grouped classrooms do learn more. This gets at the heart of the matter, mentioned at the outset: There are important factors that influence children’s learning that are not genetically built-in, but instead have social origins. What and how much students are taught varies due to shifts in things from national policies to micro-level classroom instructional practices, and such variations influence students’ learning.

While abandoning skill grouping may appear to be a logical way to reduce inequality in learning, it would also involve a trade-off: Highly-skilled students would probably learn less than if they were grouped, and low-skilled students would probably learn more than if they were grouped. Parents of highly-skilled students would likely disagree with this kind of change, whereas parents of low-skilled students would likely welcome it since their children would benefit. Such dilemmas will always be at the center of debates over curriculum differentiation, which are shaped by individuals’ overall views of the education system’s role in society.

For some, inequality in learning outcomes is expected and is perfectly acceptable. Curriculum differentiation should be retained because highly-skilled students should not lose opportunities so that low-skilled students can catch up. Society needs the education system to promote the development of the best and brightest students so they can go on to invent new technologies and discover cures for diseases. From this perspective, society might actually be worse off in the long run without curriculum differentiation performing this sifting and sorting function. For others, the education system should be working to
reduce inequality and equalize opportunities. The system of public education was founded on these principles, as Horace Mann viewed schools as “the great equalizer.” Without a common school system promoting equal opportunities for all, society would have fewer educated people from whom to benefit and there would be a greater gulf between the “haves” and the “have nots.” From this perspective, abandoning curriculum differentiation is a step toward the ideal of equal opportunity.

**Limitations and Implications for Future Research**

While this study makes an important contribution to the literature and informs theoretical and policy debates in key ways, it is not without limitations. Its strength lies in its use of a large, representative data set to gain a better understanding of skill grouping and its effects on learning on a national level, but there are limits to studying skill grouping with ECLS-K.

The primary drawback is that the students were sampled within schools, not classrooms. As noted in Chapter 6, students are scattered across classrooms within schools (especially in the first-grade wave), and this prevents researchers from analyzing skill grouping – a within-classroom process – at the classroom level. For instance, it is not possible to determine the size, homogeneity, and socioeconomic/racial composition of each skill group, the instructional methods/materials used in each group, or the overall variance in skills for the entire classroom – all things that smaller-scale studies have found to be salient (Barr and Dreeben 1983; Gamoran 1986; Hallinan and Sorensen 1983). The actual differentiation of instruction across reading groups is especially important, as the whole point of skill grouping is to vary the pace and difficulty of
reading lessons according to students’ skills. Unfortunately, the most useful information available about students’ group placement is which group they are in. If information on the differentiation of instruction was available, its effects on reading gains might prove to be even more pronounced than my results suggest.

This sampling limitation is especially relevant in Chapter 6. The analysis comparing the variance in skills of grouped and non-grouped students is not fruitful, and the analyses of socioeconomic and racial disparities in reading gains do not reveal clear evidence that skill grouping is partially to blame. I suspect that the inability to explore these hypotheses at the classroom level is primarily to blame. The main models predicting reading gains (Tables 4 and 5) test for the overall effects of students’ relative skill group placements on learning, net of a wide variety of covariates, among students across the entire sample. Those analyses are arguably the most stringent tests of whether grouping matters yet conducted. The questions pertaining to variance in skills and socioeconomic/racial gaps, however, would probably be better addressed with more detailed data. How might researchers draw the ideal sample for studying these issues?

National scope and generalizability should be a goal, but one key difference from ECLS-K would be sampling classrooms instead of schools, and then collecting information on all students in the various classrooms. Such data would allow researchers to measure the kinds of things mentioned above that cannot be measured using ECLS-K. This might not be the ideal sample for studying school-level processes, of course, but it would allow for excellent analyses of skill grouping and other instructional practices (i.e., educational triage) that occur in the context most proximate to students – the classroom.
In the end, a data set that combines the benefits of small-scale studies with the advantage of representativeness would provide researchers tremendous leverage for studying skill grouping.

A second drawback is that, as discussed in Appendix B, it appears that ECLS-K’s cognitive assessments may not have adequately measured the most advanced students’ skills. This creates challenges when studying skill grouping, because high-grouped students are the most likely to be approaching the test-score ceiling. In models of reading gains in first grade, students’ initial skills are negatively associated with gains. Because of this association, not controlling for initial skills underestimates the effects of group placement on gains – especially for high-grouped students (results not shown). Moreover, this ceiling effect poses problems for analyses of variance such as those presented in Figure 4. When the full range of students’ skills is not measured, it is difficult to draw conclusions pertaining to issues of variance. The ceiling effect could also have an impact on the analyses of socioeconomic and racial disparities in learning (Chapter 6). It could be that growth in disadvantaged students’ skills is better measured than that of their advantaged peers who are near the ceiling. If so, the impact of skill grouping on socioeconomic and racial disparities in reading gains may be unclear. I attempted to address these concerns by controlling for initial skills and by measuring and modeling percentage of maximum possible gain, but the ceiling effect might still be muddling these analyses.

Measuring and modeling learning continues to be an important challenge in the study of educational stratification, and future research should work toward improving the
methods currently being used. IRT methods have been used for decades, but researchers should not assume that the use of IRT automatically eliminates ceiling effects. As Figure 7 (discussed in Appendix B) illustrates, there is a limit to how many points ECLS-K students could have gained beyond their initial scores, and this limit appears to have influenced the relationship between initial skills and gains. For students scoring in the low and middle range at Time 1, there appears to be no relationship between initial score and gains. For students scoring high at Time 1, there is a clear negative relationship between initial score and gains. Future studies should attempt to avoid such patterns by allowing initially high-skilled students more room to exhibit gains in skills from Time 1 to Time 2. This might be accomplished by testing students’ mastery of far more advanced skills, making the test-score ceiling out of reach for even the brightest students. Only then will concerns with ceiling effects subside.

In future studies of homogeneous grouping and unequal learning outcomes specifically, it will be necessary to consider both skill grouping and educational triage practices together. Rather than asking teachers only about students’ skill group placements, researchers will also need to investigate how students fit into the triage-related efforts to raise proficiency test scores and whether such efforts overlap with traditional skill grouping strategies. It could be that in first and second grade, when there are no proficiency tests, the pace and difficulty of instruction in the low, middle, and high groups will produce patterns similar to those found in this study. Later, during tested years, students in the middle group might essentially be the “bubble kids” and exhibit the most gains of all due to the extra efforts to push them to the proficiency bar. Such
questions are for future studies, but for now the lesson is that researchers must keep up as instructional practices change. Importantly, as research on skill grouping suggests that more/better instruction is needed in the lower-ranked groups, educational triage as a response to No Child Left Behind may be a barrier to such a change.

Another topic for future research is the difference between grouped and non-grouped instruction. Roughly 30% of ECLS-K first graders were in classrooms where the teacher reported not using skill groups. First, is the use of skill grouping declining? Past research has found skill grouping for reading instruction within elementary classrooms to be so common that studies have not been able to make comparisons to non-grouped students. The ECLS-K data suggest that today there are teachers who do not use skill grouping. Past studies may have (a) focused only on grouped students intentionally or (b) not drawn samples large enough to capture many non-grouped classrooms. Either way, my study suggests that researchers addressing these issues in the future will find it fruitful to compare grouped students to similar non-grouped students.

A second question is, what factors shape a teacher’s decision whether to use skill grouping in the first place? My analyses did not reveal any systematic trends, but they did hint at the possibility that grouping is used more often as the heterogeneity of the student body increases. Table 3, for instance, suggests that grouped students tend to be in schools and classrooms with higher proportions of racial minority students, larger schools, and public schools. Moreover, grouped students’ teachers appear to be slightly younger and more educated, and Figure 4 suggests more variance in grouped students’ skills at the beginning of the school year. It makes sense that teachers with more
heterogeneous student bodies are more likely to group, but future studies should ask teachers directly what guides their decision. I suspect that teachers use homogeneous grouping not only as a way to manage heterogeneity in students’ skills, but also because small groups are more manageable than one large group. Whatever might occur in the way of policy attempts to regulate the use of skill grouping, teachers’ preferences will undoubtedly be relevant.

**Conclusion**

While there are many questions surrounding the use of homogeneous skill grouping, the primary questions for sociologists – the questions on which I have focused in this study – pertain to skill grouping’s role in the link between schooling and inequality. When it comes to stratification, few things are as relevant as an individual’s educational opportunities and outcomes. While schooling may reduce inequality in cognitive skills compared to that which emerges when school is not in session (Downey et al. 2004), considerable disparities in skills also emerge during the school year.

The research reported here suggests that early elementary, within-classroom skill grouping contributes to such disparities in learning. When students begin school, most of them are placed into a particular group for reading lessons based on their skill level. These group placements confer advantages upon some (accelerated learning for high-grouped students) and disadvantages upon others (decelerated learning for low-grouped students) compared to full classroom instruction. Moreover, this occurs during a “critical period” for young children as they learn basic literacy skills that are the foundation for future learning (Entwisle and Alexander 1993). These early educational experiences are
crucial to our broader understanding of educational attainment, stratification, and institutional processes through which inequality is reproduced. For scholars of educational stratification, as well as parents and educators concerned about inequality, skill grouping within elementary classrooms should rank high among school-related practices considered to promote disparities in skills.
LIST OF REFERENCES


Seltzer, Michael, Kilchan Choi, and Yeow Meng Thum. 2003. “Examining Relationships Between Where Students Start and How Rapidly they Progress: Using New Developments in Growth Modeling to Gain Insight into the


APPENDIX A

TABLES AND FIGURES
<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Min.</th>
<th>Max.</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading gains</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall reading skills</td>
<td>12.57</td>
<td>86.63</td>
<td>38.96</td>
<td>12.77</td>
</tr>
<tr>
<td>Reading gains</td>
<td>-21.88</td>
<td>46.05</td>
<td>17.52</td>
<td>8.05</td>
</tr>
<tr>
<td>Percentage of maximum possible gain</td>
<td>-37.41</td>
<td>82.72</td>
<td>34.14</td>
<td>15.63</td>
</tr>
<tr>
<td>Skill grouping</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill group rank</td>
<td>1.25</td>
<td>8.75</td>
<td>5.46</td>
<td>2.22</td>
</tr>
<tr>
<td>Grouped</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.68</td>
<td></td>
</tr>
<tr>
<td>Lowest skill group</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>Middle skill group</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.26</td>
<td></td>
</tr>
<tr>
<td>Highest skill group</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.24</td>
<td></td>
</tr>
<tr>
<td>Non-grouped</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.32</td>
<td></td>
</tr>
<tr>
<td>Other measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest SES quintile</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.23</td>
<td></td>
</tr>
<tr>
<td>Fourth SES quintile</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.20</td>
<td></td>
</tr>
<tr>
<td>Third SES quintile</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.20</td>
<td></td>
</tr>
<tr>
<td>Second SES quintile</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.19</td>
<td></td>
</tr>
<tr>
<td>Lowest SES quintile</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.56</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.17</td>
<td></td>
</tr>
<tr>
<td>Other race</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>Both biological parents in home</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.65</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.50</td>
<td></td>
</tr>
<tr>
<td>Age (in months in fall of school year)</td>
<td>67.97</td>
<td>98.73</td>
<td>80.10</td>
<td>4.47</td>
</tr>
<tr>
<td>Student has disability</td>
<td>0 (no)</td>
<td>1 (yes)</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>Approaches to learning</td>
<td>1</td>
<td>4</td>
<td>3.04</td>
<td>.69</td>
</tr>
<tr>
<td>Externalizing problem behaviors</td>
<td>1</td>
<td>4</td>
<td>1.68</td>
<td>.62</td>
</tr>
<tr>
<td>Number of days between student’s fall and spring assessments</td>
<td>146</td>
<td>281</td>
<td>208.45</td>
<td>20.81</td>
</tr>
<tr>
<td>School’s percentage of economically disadvantaged students</td>
<td>0</td>
<td>95</td>
<td>29.39</td>
<td>25.80</td>
</tr>
<tr>
<td>School’s percentage of racial minority students</td>
<td>1</td>
<td>5</td>
<td>2.80</td>
<td>1.56</td>
</tr>
<tr>
<td>School’s total enrollment</td>
<td>1</td>
<td>5</td>
<td>3.43</td>
<td>1.18</td>
</tr>
</tbody>
</table>

Table 1. Descriptive Statistics for Measures Used in the Study
Table 1 continued

<table>
<thead>
<tr>
<th></th>
<th>Minimums</th>
<th>Maximums</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private school</td>
<td>0 (no)</td>
<td>1 (yes)</td>
</tr>
<tr>
<td>Non-White teacher</td>
<td>0 (no)</td>
<td>1 (yes)</td>
</tr>
<tr>
<td>Teacher’s age</td>
<td>23</td>
<td>65</td>
</tr>
<tr>
<td>Teacher’s level of education</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Teacher’s number of years of experience</td>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>Classroom’s total number of students</td>
<td>12</td>
<td>35</td>
</tr>
<tr>
<td>Classroom’s percentage of racial minority students</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Classroom’s adequacy of textbooks</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Adequacy of classroom space</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Teacher’s report of students’ behavior</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

Minimums and maximums are observed values in the original subsample of first graders; means and standard deviations are derived from one of the five data sets with imputed missing values.
### Table 2. Generalized Least Squares Regression Coefficients for Hypothesized Predictors of Skill Group Rank

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Socioeconomic Status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fourth SES quintile (vs. highest)</td>
<td>-0.57*** (.14)</td>
<td>-0.54*** (.14)</td>
<td>0.06 (.12)</td>
<td></td>
</tr>
<tr>
<td>Third SES quintile (vs. highest)</td>
<td>-0.72*** (.14)</td>
<td>-0.62*** (.14)</td>
<td>0.13 (.13)</td>
<td></td>
</tr>
<tr>
<td>Second SES quintile (vs. highest)</td>
<td>-0.96*** (.16)</td>
<td>-0.81*** (.16)</td>
<td>0.16 (.13)</td>
<td></td>
</tr>
<tr>
<td>Lowest SES quintile (vs. highest)</td>
<td>-1.55*** (.18)</td>
<td>-1.44*** (.20)</td>
<td>0.14 (.15)</td>
<td></td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian (vs. White)</td>
<td>0.45 (.30)</td>
<td>0.40 (.30)</td>
<td>-0.15 (.27)</td>
<td></td>
</tr>
<tr>
<td>Black (vs. White)</td>
<td>-0.67*** (.17)</td>
<td>-0.28 (.18)</td>
<td>-0.06 (.16)</td>
<td></td>
</tr>
<tr>
<td>Hispanic (vs. White)</td>
<td>-0.44* (.18)</td>
<td>-0.12 (.18)</td>
<td>0.19 (.16)</td>
<td></td>
</tr>
<tr>
<td>Other race (vs. White)</td>
<td>-0.23 (.20)</td>
<td>0.09 (.20)</td>
<td>0.30 (.19)</td>
<td></td>
</tr>
<tr>
<td><strong>Background Controls</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both biological parents in home (vs. all other families)</td>
<td>0.54*** (.12)</td>
<td>0.26* (.11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (vs. male)</td>
<td>0.61*** (.09)</td>
<td>0.01 (.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.04*** (.01)</td>
<td>-0.02* (.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student has disability (vs. does not)</td>
<td>-0.80*** (.15)</td>
<td>-0.03 (.11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>School Composition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of economically disadvantaged students</td>
<td>0.01 (.00)</td>
<td>0.00 (.00)</td>
<td>0.01* (.00)</td>
<td></td>
</tr>
<tr>
<td>Percentage of racial minority students</td>
<td>0.10* (.05)</td>
<td>0.15*** (.05)</td>
<td>0.09* (.05)</td>
<td></td>
</tr>
<tr>
<td><strong>Effort, Behavior, and Initial Skills</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approaches to learning</td>
<td>1.37*** (.09)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Externalizing problem behaviors</td>
<td></td>
<td>0.38*** (.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall reading skills</td>
<td></td>
<td>0.09*** (.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intercept</strong></td>
<td>5.90</td>
<td>5.23</td>
<td>1.73</td>
<td>-2.27</td>
</tr>
<tr>
<td>R²</td>
<td>0.04</td>
<td>0.01</td>
<td>0.08</td>
<td>0.42</td>
</tr>
</tbody>
</table>

* p < .05; ** p < .01; *** p < .001 (two-tailed tests); Standard errors are in parentheses; N = 3,363 grouped students in 521 schools and 2 strata.
Table 3. Results of T-Tests Comparing Means of Selected Characteristics of Grouped and Non-grouped Students’ Schools, Classrooms, and Teachers

<table>
<thead>
<tr>
<th></th>
<th>Grouped Students (N=3,374)</th>
<th>Sig. of t-test</th>
<th>Non-grouped Students (N=1,579)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of economically disadvantaged students</td>
<td>28.71</td>
<td>**</td>
<td>30.86</td>
</tr>
<tr>
<td>Percentage of racial minority students</td>
<td>2.84</td>
<td>**</td>
<td>2.69</td>
</tr>
<tr>
<td>Total enrollment</td>
<td>3.52</td>
<td>***</td>
<td>3.22</td>
</tr>
<tr>
<td>Private (vs. public)</td>
<td>.18</td>
<td>***</td>
<td>.27</td>
</tr>
<tr>
<td><strong>Classroom Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of students</td>
<td>20.67</td>
<td>n/s</td>
<td>20.70</td>
</tr>
<tr>
<td>Percentage of racial minority students</td>
<td>36.67</td>
<td>***</td>
<td>32.90</td>
</tr>
<tr>
<td>Adequacy of textbooks</td>
<td>3.36</td>
<td>n/s</td>
<td>3.38</td>
</tr>
<tr>
<td>Adequacy of classroom space</td>
<td>3.09</td>
<td>n/s</td>
<td>3.07</td>
</tr>
<tr>
<td>Teacher’s report of students’ behavior</td>
<td>3.40</td>
<td>n/s</td>
<td>3.42</td>
</tr>
<tr>
<td><strong>Teacher Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-White</td>
<td>.13</td>
<td>n/s</td>
<td>.12</td>
</tr>
<tr>
<td>Age</td>
<td>42.90</td>
<td>**</td>
<td>43.81</td>
</tr>
<tr>
<td>Level of education</td>
<td>3.16</td>
<td>***</td>
<td>3.06</td>
</tr>
<tr>
<td>Number of years of experience</td>
<td>14.71</td>
<td>n/s</td>
<td>14.96</td>
</tr>
</tbody>
</table>

* p < .05; ** p < .01; *** p < .001 (two-tailed t-tests); Analyses based on one of the five data sets with imputed missing values.
<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Reading Gains</th>
<th>Percentage of Max. Poss. Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>Grouped (vs. non-grouped)</td>
<td>-22</td>
<td>-.37</td>
</tr>
<tr>
<td>Lowest skill group (vs. non-grouped)</td>
<td>-2.73***</td>
<td>-5.00***</td>
</tr>
<tr>
<td>Middle skill group (vs. non-grouped)</td>
<td>.17</td>
<td>-.47</td>
</tr>
<tr>
<td>Highest skill group (vs. non-grouped)</td>
<td>1.25**</td>
<td>3.07**</td>
</tr>
<tr>
<td>Fall reading skills</td>
<td>-26***</td>
<td>-29***</td>
</tr>
<tr>
<td>Approaches to learning</td>
<td>3.59***</td>
<td>3.08***</td>
</tr>
<tr>
<td>Externalizing problem behaviors</td>
<td>.09</td>
<td>-.06</td>
</tr>
<tr>
<td>Both biological parents in home (vs. all other families)</td>
<td>-.23</td>
<td>-.31</td>
</tr>
<tr>
<td>Female (vs. male)</td>
<td>-.52*</td>
<td>-.55*</td>
</tr>
<tr>
<td>Age</td>
<td>-1.17***</td>
<td>-1.16***</td>
</tr>
<tr>
<td>Student has disability (vs. does not)</td>
<td>-1.083**</td>
<td>-1.97**</td>
</tr>
<tr>
<td>Fourth SES quintile (vs. highest)</td>
<td>.26</td>
<td>.26</td>
</tr>
<tr>
<td>Third SES quintile (vs. highest)</td>
<td>-.45</td>
<td>-.45</td>
</tr>
<tr>
<td>Second SES quintile (vs. highest)</td>
<td>-.75</td>
<td>-.73</td>
</tr>
<tr>
<td>Lowest SES quintile (vs. highest)</td>
<td>-.99</td>
<td>-.93</td>
</tr>
<tr>
<td>Asian (vs. White)</td>
<td>-1.54***</td>
<td>-1.87***</td>
</tr>
<tr>
<td>Black (vs. White)</td>
<td>-.59</td>
<td>-.53</td>
</tr>
<tr>
<td>Hispanic (vs. White)</td>
<td>-.59</td>
<td>-.65</td>
</tr>
<tr>
<td>Other race (vs. White)</td>
<td>-.84</td>
<td>-.104</td>
</tr>
<tr>
<td>Classroom’s percentage of racial minority students</td>
<td>-.01</td>
<td>-.01</td>
</tr>
</tbody>
</table>

Table 4. Generalized Least Squares Regression Coefficients Estimating the Effects of Skill Group Placement on Reading Gains
Table 4 continued

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate 1</th>
<th>Estimate 2</th>
<th>Estimate 3</th>
<th>Estimate 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher’s level of education</td>
<td>.16</td>
<td>.22</td>
<td>.17</td>
<td>.30</td>
</tr>
<tr>
<td></td>
<td>(.17)</td>
<td>(.18)</td>
<td>(.33)</td>
<td>(.34)</td>
</tr>
<tr>
<td>Teacher’s age</td>
<td>.05**</td>
<td>.04**</td>
<td>.10***</td>
<td>.09**</td>
</tr>
<tr>
<td></td>
<td>(.01)</td>
<td>(.01)</td>
<td>(.03)</td>
<td>(.03)</td>
</tr>
<tr>
<td>School’s percentage of econ. disadvantaged</td>
<td>-.01</td>
<td>-.01</td>
<td>-.03*</td>
<td>-.03*</td>
</tr>
<tr>
<td>students</td>
<td>(.01)</td>
<td>(.01)</td>
<td>(.01)</td>
<td>(.01)</td>
</tr>
<tr>
<td>School’s percentage of racial minority students</td>
<td>-.45*</td>
<td>-.43*</td>
<td>-.87*</td>
<td>-.84*</td>
</tr>
<tr>
<td></td>
<td>(.19)</td>
<td>(.19)</td>
<td>(.36)</td>
<td>(.36)</td>
</tr>
<tr>
<td>School’s total enrollment</td>
<td>-.18</td>
<td>-.21</td>
<td>-.35</td>
<td>-.43</td>
</tr>
<tr>
<td></td>
<td>(.15)</td>
<td>(.15)</td>
<td>(.29)</td>
<td>(.29)</td>
</tr>
<tr>
<td>Private school (vs. public)</td>
<td>.87</td>
<td>1.07*</td>
<td>2.00</td>
<td>2.22*</td>
</tr>
<tr>
<td></td>
<td>(.53)</td>
<td>(.53)</td>
<td>(1.06)</td>
<td>(1.05)</td>
</tr>
<tr>
<td>Number of days between assessments</td>
<td>.05***</td>
<td>.05****</td>
<td>.10***</td>
<td>.10****</td>
</tr>
<tr>
<td></td>
<td>(.01)</td>
<td>(.01)</td>
<td>(.01)</td>
<td>(.01)</td>
</tr>
<tr>
<td>Intercept</td>
<td>19.71</td>
<td>22.52</td>
<td>10.34</td>
<td>16.46</td>
</tr>
<tr>
<td>$R^2$</td>
<td>20</td>
<td>22</td>
<td>21</td>
<td>23</td>
</tr>
</tbody>
</table>

* $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed tests); Standard errors are in parentheses; N = 4,953 students in 668 schools and 2 strata.
<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grouped (vs. non-grouped)</td>
<td>.27 (.28)</td>
<td>.74 (.58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lowest skill group (vs. non-grouped)</td>
<td></td>
<td>-1.22* (.52)</td>
<td>-1.67 (1.05)</td>
<td></td>
</tr>
<tr>
<td>Middle skill group (vs. non-grouped)</td>
<td>.76 (.45)</td>
<td>1.32 (.91)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest skill group (vs. non-grouped)</td>
<td></td>
<td>.91* (.44)</td>
<td>2.12* (.88)</td>
<td></td>
</tr>
<tr>
<td>Fall reading skills</td>
<td>-.78*** (.03)</td>
<td>-.79*** (.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time: Kindergarten (vs. first grade)</td>
<td>-18.72*** (.58)</td>
<td>-18.95*** (.64)</td>
<td>-17.65*** (.72)</td>
<td>-17.77*** (.75)</td>
</tr>
<tr>
<td>Number of days between assessments</td>
<td>.04*** (.01)</td>
<td>.04*** (.01)</td>
<td>.09*** (.02)</td>
<td>.09*** (.02)</td>
</tr>
<tr>
<td>Within-student R²</td>
<td>.57</td>
<td>.57</td>
<td>.56</td>
<td>.57</td>
</tr>
<tr>
<td>Overall R²</td>
<td>.12</td>
<td>.12</td>
<td>.34</td>
<td>.37</td>
</tr>
<tr>
<td>N (student-years)</td>
<td>3,201</td>
<td>2,855</td>
<td>3,201</td>
<td>2,855</td>
</tr>
</tbody>
</table>

* p < .05; ** p < .01; *** p < .001 (two-tailed tests); Standard errors are in parentheses

Table 5. Results of Cross-Sectional Time-Series Analyses with Fixed Effects Estimating Within-Student Reading Gains Across Kindergarten and First Grade
<table>
<thead>
<tr>
<th></th>
<th>Low-High SES Gap</th>
<th>Black-White Gap</th>
<th>Hispanic-White Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grouped students</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td>15.38</td>
<td>5.46</td>
<td>6.97</td>
</tr>
<tr>
<td>Spring</td>
<td>17.12</td>
<td>7.48</td>
<td>7.26</td>
</tr>
<tr>
<td>Increase</td>
<td>1.74</td>
<td>2.02</td>
<td>.29</td>
</tr>
<tr>
<td><strong>Non-grouped students</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td>13.37</td>
<td>5.75</td>
<td>6.91</td>
</tr>
<tr>
<td>Spring</td>
<td>15.07</td>
<td>8.46</td>
<td>8.77</td>
</tr>
<tr>
<td>Increase</td>
<td>1.70</td>
<td>2.71</td>
<td>1.86</td>
</tr>
</tbody>
</table>

Table 6. Growth in Socioeconomic and Racial Gaps in Reading Skills Among Grouped and Non-Grouped Students
<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fourth SES quintile (vs. highest)</td>
<td>-.20</td>
<td>-.21</td>
<td>-1.89*</td>
<td>-1.34</td>
</tr>
<tr>
<td></td>
<td>(.37)</td>
<td>(.36)</td>
<td>(.76)</td>
<td>(.74)</td>
</tr>
<tr>
<td>Third SES quintile (vs. highest)</td>
<td>-1.49***</td>
<td>-1.56***</td>
<td>-5.32***</td>
<td>-4.61***</td>
</tr>
<tr>
<td></td>
<td>(.41)</td>
<td>(.40)</td>
<td>(.77)</td>
<td>(.79)</td>
</tr>
<tr>
<td>Second SES quintile (vs. highest)</td>
<td>-2.02***</td>
<td>-2.14***</td>
<td>-6.68***</td>
<td>-5.85***</td>
</tr>
<tr>
<td></td>
<td>(.44)</td>
<td>(.44)</td>
<td>(.84)</td>
<td>(.86)</td>
</tr>
<tr>
<td>Lowest SES quintile (vs. highest)</td>
<td>-4.24***</td>
<td>-4.27***</td>
<td>-11.57***</td>
<td>-9.98***</td>
</tr>
<tr>
<td></td>
<td>(.51)</td>
<td>(.50)</td>
<td>(.94)</td>
<td>(.92)</td>
</tr>
<tr>
<td>Number of days between assessments</td>
<td>.05***</td>
<td>.05***</td>
<td>.09***</td>
<td>.09***</td>
</tr>
<tr>
<td></td>
<td>(.01)</td>
<td>(.01)</td>
<td>(.02)</td>
<td>(.02)</td>
</tr>
<tr>
<td>Fall reading skills</td>
<td>-.17***</td>
<td>-.22***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.01)</td>
<td>(.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lowest skill group (vs. non-grouped)</td>
<td>-3.79***</td>
<td></td>
<td>-7.84***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.51)</td>
<td></td>
<td>(.94)</td>
<td></td>
</tr>
<tr>
<td>Middle skill group (vs. non-grouped)</td>
<td>-.15</td>
<td></td>
<td>-1.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.45)</td>
<td></td>
<td>(.87)</td>
<td></td>
</tr>
<tr>
<td>Highest skill group (vs. non-grouped)</td>
<td>1.30**</td>
<td></td>
<td>4.16***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.45)</td>
<td></td>
<td>(.95)</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>14.90</td>
<td>17.70</td>
<td>19.93</td>
<td>20.28</td>
</tr>
<tr>
<td>R²</td>
<td>.09</td>
<td>.13</td>
<td>.07</td>
<td>.13</td>
</tr>
</tbody>
</table>

* p < .05; ** p < .01; *** p < .001 (two-tailed tests); Standard errors are in parentheses; N=4,953 students in 668 schools and 2 strata

Table 7. Results of Generalized Least Squares Regressions Testing for Mediation of Socioeconomic Gaps in Reading Gains
<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Reading Gains</th>
<th>Percentage of Max. Poss. Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>Asian (vs. White)</td>
<td>-1.93**</td>
<td>-2.05***</td>
</tr>
<tr>
<td></td>
<td>(.59)</td>
<td>(.59)</td>
</tr>
<tr>
<td>Black (vs. White)</td>
<td>-3.09***</td>
<td>-3.14***</td>
</tr>
<tr>
<td></td>
<td>(.40)</td>
<td>(.39)</td>
</tr>
<tr>
<td>Hispanic (vs. White)</td>
<td>-2.61***</td>
<td>-2.86***</td>
</tr>
<tr>
<td></td>
<td>(.45)</td>
<td>(.46)</td>
</tr>
<tr>
<td>Other race (vs. White)</td>
<td>-2.06***</td>
<td>-2.37***</td>
</tr>
<tr>
<td></td>
<td>(.57)</td>
<td>(.58)</td>
</tr>
<tr>
<td>Number of days between</td>
<td></td>
<td></td>
</tr>
<tr>
<td>assessments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall reading skills</td>
<td>-.15***</td>
<td>-.20***</td>
</tr>
<tr>
<td></td>
<td>(.01)</td>
<td>(.01)</td>
</tr>
<tr>
<td>Lowest skill group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(vs. non-grouped)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle skill group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(vs. non-grouped)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest skill group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(vs. non-grouped)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>13.19</td>
<td>15.93</td>
</tr>
<tr>
<td>R²</td>
<td>.09</td>
<td>.13</td>
</tr>
</tbody>
</table>

* p < .05; ** p < .01; *** p < .001 (two-tailed tests); Standard errors are in parentheses; N=4,953 students in 668 schools and 2 strata

Table 8. Results of Generalized Least Squares Regressions Testing for Mediation of Racial Gaps in Reading Gains
Figure 1. Hypothesized Effects of Skill Grouping on Reading Gains
*** $p < .001$ (two-tailed t-tests comparing skill group rank to that of highest SES quintile)

Figure 2. Socioeconomic Disparities in Skill Group Placement
5.23
4.56***
4.79*

*** p < .001; ** p < .01; * p < .05 (two-tailed t-tests comparing skill group rank to that of White students)

Figure 3. Racial Disparities in Skill Group Placement
Fall, Non-grouped  Fall, Grouped  Spring, Non-grouped  Spring, Grouped
Mean  38.75  39.06  56.82  56.73

* $p < .05$; two-tailed test indicating unequal variances for grouped and non-grouped students. Bars represent mean plus and minus one standard deviation.

Figure 4. Variance in Reading Skills in the Fall and Spring of First Grade for Grouped and Non-Grouped Students
Figure 5. Change in Variance in Reading Skills from Fall to Spring by Skill Group Status
Figure 6. Scatter Plot of Fall and Spring Reading IRT Scores
Figure 7. Scatter Plot of Fall Reading IRT Score and Gains During First Grade
APPENDIX B

MEASURING AND MODELING LEARNING
APPENDIX B: MEASURING AND MODELING LEARNING

In this study, the dependent variables are measures of students’ progress in learning to read over the course of the school year. The first, *reading gains*, is simply how many points higher on the IRT scale the students scored in the spring of the school year than in the fall (y2–y1). The second, *percentage of maximum possible gain*, expresses students’ gains as a percentage of how many points they could have possibly gained (((y2–y1)/(92–y1))*100). When predicting reading gains, y1 is included as a control variable; when predicting percentage of maximum possible gain, y1 is *not* included as a control. The purpose of this Appendix is to explain these interrelated decisions pertaining to the measurement and modeling of learning using the ECLS-K data.

The first point to be made is that predicting *gains* is a more favorable way to gauge *learning* than is predicting a test score at one point in time. “Proficiency tests,” for example, are mandated by state and federal laws, and performance is evaluated relative to cut-off points that indicate various levels of mastery (i.e., below average, proficient, above average, accelerated, etc.). When researchers explore the determinants of test scores, the equation looks something like the following:

\[ y = a + b_1x_1 + b_2x_2 + \ldots + b_nx_n + e \quad (1) \]

where *y* is the test score, *a* is the intercept, there are *n* independent variables, *x*, with *n* slopes, *b*, and an error term, *e*. I refer to Equation 1 as the “test score model.”
While a test score from one point in time certainly tells us something about how much a student knows, it tells us little or nothing about how much a student learns during the school year. For an analysis of whether a school process shapes learning, such as that being undertaken here, two test scores are needed – one before and one after exposure to the independent variable. It is only in this way that researchers can assess whether the independent variable has an effect on gains in skills over time. One approach is to control for $y_1$ (a measure of the dependent variable at Time 1) when predicting $y_2$ (a measure of the dependent variable at Time 2), which I refer to as the “lagged test score model:”

$$y_2 = a + b_0y_{10} + b_1x_1 + b_2x_2 + \ldots b_nx_n + e$$  \hspace{1cm} (2)

By including a control for $y_1$ on the right-hand side of the equation, the lagged test score model estimates the effects of the other independent variables on $y_2$ net of students’ scores on $y_1$; in other words, it estimates gains made between Time 1 and Time 2. Importantly, the control for $y_1$ does not fully account for differences between groups with different values on $y_1$ (Allison 1990); it is necessary to also control for other student attributes in the effort to statistically equalize students and isolate the effects of the key test variables.

As noted in the text, with IRT scale scores at two points in time (like those calculated by ECLS-K and used in this study), it is possible to measure and predict actual gains, or change scores. In the second approach to modeling gains, which I refer to as the “gain model,” the right-hand side of the equation is the same as that in Equation 1, but the dependent variable is now the difference between the fall and spring test scores:
\[ y_2 - y_1 = a + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + e \quad (3) \]

Researchers have noted two major concerns with measuring and modeling gains, or change scores, in this way: Change scores are less reliable than the two variables that make them up, and regression toward the mean can bias estimates of gains. That is, due to measurement error, initially high-skilled students’ scores are more likely to slip downward at Time 2 while initially low-skilled students’ scores are more likely to shift upward at Time 2 (Allison 1990; Willett 1988).

Some have made the case that including \( y_1 \) as a control variable lessens those concerns (Allison 1990; Morgan and Sorensen 1999; Sorensen and Morgan 2000):

\[ y_2 - y_1 = a + b_0 y_1 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + e \quad (4) \]

In the “lagged gain model” (which others call the “regressor variable model”), the right-hand side of the equation is the same as in the lagged test score model (there is a control for \( y_1 \)), while the left-hand side of the equation is the same as in the gain model (the dependent variable is gains made between Time 1 and Time 2). The lagged gain model and the lagged test score model produce identical estimates of the effects of all the independent variables except for \( y_1 \). In the lagged test score model, the coefficient for the effect of \( y_1 \) is the coefficient from lagged gain model plus 1 (see Morgan 2001; Sorensen and Morgan 2000). Importantly, “[t]hese formulations, with [\( y_1 \)] on the right-hand side, are usually seen as preferable to other models because they avoid the problems caused by the well-known unreliability of the change score, [\( y_2 - y_1 \)], and they capture the regression toward the mean that otherwise may seriously bias other coefficients correlated with [\( y_1 \)]” (Sorensen and Morgan 2000:157).
Among the models discussed here, Models (2) and (4) are the most commonly used in the current sociology of education literature. Moreover, a negative association between initial skills and gains is evident in many of these studies. Morgan and Sorensen (1999), for instance, used the lagged gain model to predict gains from 10th to 12th grades using the National Education Longitudinal Study (NELS). In an appendix, they discussed the differences between the lagged gain model and the gain model, reporting that the two approaches produced “remarkably similar” results. Still, there was a significant negative association between y1 and y2–y1 in their lagged gain model. Blau et al. (2001) used a two-subject composite as a Time-1 control in their lagged gain model but still found a significant negative association between initial skills and gains. In a recent study of summer learning using the ECLS-K data, Burkam et al. (2004) chose the lagged gain model and also noted a significant negative association between y1 and y2–y1. Finally, even researchers using multi-level growth models have estimated the effects of initial score on subsequent gains (Downey et al. 2004; Seltzer, Choi, and Thum 2003).

To test for effects of skill group placement on learning to read, I use the lagged gain model for three key reasons. First, it allows me to predict gains during the school year while lessening the potential for regression toward the mean to bias estimates (noted above). Second, the control for y1 is necessary from a substantive standpoint because initial skills are associated with both skill group placement (the key independent variable) and reading gains, the dependent variable. Not including y1 as a control, then, would be a classic case of omitted variable bias. Third, there is evidence that ECLS-K
might not have been entirely successful in its use of IRT methods to avoid ceiling effects in the test-score data. If this is the case, then students scoring higher in the fall had, by default, less room to move up the scale than their lower-skilled counterparts. Controlling for initial skills captures this phenomena if it is indeed a reality.

Still, there is a significant negative association between initial skills and gains in the first-grade school-year models presented in this study (see Table 4, Models 1 and 2). This could be due to either (a) regression toward the mean from Time 1 to Time 2, (b) a reality in which students who begin the school year with more reading skills actually learn less than their lower-skilled counterparts, or (c) a ceiling effect in the test-score data. Morgan and Sorensen (1999:668) alluded to two of these possibilities in their study of Catholic and public schools using NELS: “In all models, the tenth-grade math score has a negative relationship with gains in math achievement between the tenth and twelfth grade. The negative coefficient indicates that there is either regression toward the mean between the tests as a result of measurement error or that math learning is governed inherently by a concave growth function.” Either of these could be true of the reading test scores in ECLS-K, but another possibility that requires further discussion is that of a ceiling effect.

There are at least three pieces of evidence suggesting a ceiling effect in the reading IRT scale that affects analyses of learning in first grade. First, inspection of the frequency distribution reveals that some students scored as high as 86 out of a possible 92 in the fall, and that roughly 10% of the students in the fall subsample scored 60 or higher in the fall. This means that many students with higher initial skill levels simply
did not have as much room on the scale to gain points relative to initially lower-skilled students. Second, a scatter plot of the spring reading test score on the fall reading test score makes it clear that this phenomenon had an impact on the relationship between fall and spring test scores. Figure 6 shows a positive linear relationship in which students with more skills in the fall tend to also have more skills in the spring. This relationship changes, however, near the top of the IRT scale – among students scoring above approximately 50 in the fall (16% of the sample), the positive relationship between fall and spring scores is visibly weaker. This is further illustrated by Figure 7, which is a scatter plot of gains on the fall test score. Figure 7 illustrates the negative relationship between initial score and gains, and it also shows that this relationship comes about mainly due to the students who scored higher than approximately 50 in the fall. Finally, ECLS-K stated in their user’s manual for the first-grade data release that there was some concern with a ceiling effect even in the spring of kindergarten: “Analysis of the reading scores from spring-kindergarten showed a higher than expected number of respondents scoring near the ceiling. Therefore, to eliminate the possibility of ceiling effects, the number of reading items was increased by adding more difficult vocabulary words and text” (National Center for Education Statistics 2002:2-6).

To summarize, ECLS-K recognized the possibility of a ceiling effect at the end of kindergarten and attempted to address this concern, but the first-grade patterns illustrated in Figures 6 and 7 suggest that this attempt was not completely successful. This affects my study directly in a couple of important ways. First, in addition to reading gains, I also calculate and predict students’ percentage of maximum possible gain. This measure “has
the potential for helping to minimize the impact of ceiling effects if they should occur”
because it takes into account the varying distances between students’ initial scores and
the maximum possible score on the scale (U.S. Department of Education 2002b:8-15).
For example, a student with a score of 86 in the fall and 89 in the spring gained 3 points,
as did a student who scored 41 in the fall and 44 in the spring. However, the former
achieved 50% of maximum possible gain while the latter only achieved 6% of maximum
possible gain. This measure assures that the student near the top of the scale gets “more
credit” for gaining 3 points at the difficult top end of the scale than the student who also
gained 3 points but made gains in the less difficult range of skills and had much more
room (more points) on the IRT scale to grow. Importantly, the models predicting
percentage of maximum possible gain produce findings consistent with the lagged gain
models, suggesting that the ceiling effect is neutralized in the multivariate analyses and
that the results from the lagged gain models are robust.

The ceiling effect also has implications for the results presented in Chapter 6.
The reproduction of inequality perspective predicts that the test-score distribution of
grouped students should grow more unequal than that of non-grouped students, but I
found that not to be the case (see Figure 4). If the top end of the reading IRT scale is
truncated by a ceiling effect, the extent of inequality in the test-score distribution is also
truncated. This, in turn, could affect the comparisons of grouped and non-grouped
students. Since the results presented in Chapters 4 and 5 suggest that initially high-
skilled students tend to be placed into high-ranked groups, and that high-grouped
students gain more than similar non-grouped students, high-grouped students are the
most likely to reach the test-score ceiling. This is certainly suggested by Figure 5, which shows that high-grouped students actually exhibit a decrease in their test-score standard deviation from fall to spring, and would explain the inconsistency between the multivariate models predicting learning (which includes a control for initial skills) and the bivariate comparisons of test-score distributions in Figures 4 and 5.
APPENDIX C

NOTES ON ROBUSTNESS OF RESULTS
APPENDIX C: NOTES ON ROBUSTNESS OF RESULTS

In the course of a project, researchers inevitably conduct several analyses in addition to those that appear in the final product. This dissertation is no exception. In this Appendix, I document the many supplemental analyses I conducted in order to ensure that the findings presented in the text are robust.

The Impact of Missing Data

As noted in the text, I chose to use multiple imputation of missing data in order to use the best possible method of maximizing my sample size. Since this procedure is not well known and may make some readers weary of the results, I ran supplemental models using listwise deletion of cases with missing values. These supplemental analyses produced results very similar to those presented in the text. In no case did an analysis using listwise deletion lead to a different conclusion than that drawn from the models using imputed data. In some cases, significance levels of some coefficients declined as listwise deletion of cases with missing values resulted in much lower N’s (especially in Table 4, where more than 20 variables are entered into the equations). Overall, the use of multiple imputation is not a factor that shapes the findings and/or conclusions of this study.
Mobility in Skill Group Placement

The ECLS-K questionnaire asked teachers whether the focal child was moved to a higher group, moved to a lower group, or remained in the same group all year. It is possible that students’ mobility from one group to another could have an impact on my results. Unfortunately, all of the questions pertaining to teachers’ use of skill grouping were asked in the spring, so it is impossible to know whether teachers’ reports of students’ skill group placements reflect their initial placement or a subsequent higher or lower placement. This muddles the measurement of skill group placement somewhat, and I addressed this limitation in supplemental analyses.

First, I excluded all grouped students whose teachers reported that they moved from one group to another during the school year (roughly one third of the grouped students) and re-ran the models presented in Table 4. In each case, the analyses with movers excluded replicate those with them included. This bolsters the conclusion that group placement influences reading gains, since the overall pattern of results is consistent among students who did not change groups. Second, I entered controls into the models in Table 4 comparing students who moved up and down to those who did not move. In these analyses, students who moved up learn more, while students who moved down learn less, compared to non-movers. The estimates of the effects of low- and high-group placement are similar.

I chose not to emphasize the issue of group mobility because the measures are too vague to allow for sound analyses. The measures do not reveal why, when, or for how long the teachers moved the student to a different group. The supplemental models
suggest that group mobility might be salient, and research using more detailed data on this topic is warranted.

**Fall First Grade vs. Spring Kindergarten as Control for Initial Skills**

ECLS-K collected information on only a subsample of students in the fall of first grade, so analyses of school-year gains in skills must be limited to about 5,000 students. It is a random subsample, and results from first-grade analyses are generalizable to first graders in the U.S. in 1999-2000. Still, researchers lose a great deal of information and statistical power when limiting analyses to an N of 5,000 when data on over 17,000 are available.

To make full use of the data and ensure that my findings can be replicated using the full sample, I re-ran the main models predicting reading gains using the spring kindergarten reading score as my Time-1 indicator of skills. This involved, first, recalculating the dependent variables by subtracting the spring kindergarten score from the spring first grade score to obtain a measure of gains. Then, I used the spring kindergarten score as the control for initial skills (when predicting reading gains). Findings from these analyses solidify those presented in Table 4 that draw on the subsample. The estimates of the effects of skill grouping are not only similar, but they are even more statistically significant due to the larger N.

I still chose to use the subsample in order to measure gains in skills that occur *during the school year* and exclude gains/losses that occur over the summer, when school is not in session. This approach allowed me to test whether a process occurring within schools has an influence on gains in skills when school is in session, and to rule out the possibility that differences in summer learning bias my results. The fact that the findings
from the supplemental models reinforce those from the models restricted to the smaller subsample ensures that the choice to use the subsample does not have an impact on the conclusions drawn.

**Indicators of Skill Grouping**

There are three ECLS-K teacher questionnaires, and questions about skill grouping appear in two of them. I drew from part C, where teachers were asked to identify the group of the focal child and how many groups they use in their class. In part A, teachers were asked whether they use skill grouping, whether the focal child moved from one group to another, and some limited questions about how much time they spend in groups. This set of questions, however, did not ask teachers to identify the group placement of the focal student. I therefore had to use the questions from part C (t4chrgdgp, t4nordgp) in order to determine students’ relative group rank, but I had choices when it comes to the dichotomous indicator of whether the student was grouped.

I chose to code my “grouped” indicator based on the part-C question of which group the teacher assigned the focal student. If the teacher marked “not applicable,” I coded the “grouped” variable 0; if the teacher reported a group placement for the focal student, I coded the “grouped” variable 1. To ensure that this decision did not influence my findings, I replicated the analyses comparing grouped and non-grouped students as a whole using a grouping indicator derived from the part-A question. The findings reported in the text hold true regardless of the indicator of the use of grouping.
Patterns in Integrated Schools

The analyses testing for larger socioeconomic and racial gaps in reading gains among grouped students did not find the expected patterns. It appears that socioeconomic and racial gaps emerge similarly regardless of grouping. One possibility suggested in the literature is that skill grouping’s role in socioeconomic and racial disparities in learning might be most evident in schools that have heterogeneous student bodies. For example, Black-White disparities in group placement and reading gains would be less pronounced, and, methodologically speaking, less detectable, in majority-Black or majority-White schools. Along these lines, Ferguson (1998:365) stated: “I suspect that grouping and tracking patterns in heavily integrated schools may have a more distinctly racial component than one sees in the aggregated data.”

To address this possibility, I replicated several analyses with the sample limited to students in integrated schools (i.e., schools with at least 10% but less than 75% minority students). These models did not reveal any patterns that differ from those presented in the text: Disparities in skill group placement are not more pronounced when the sample is limited to integrated schools, the effects of skill group rank on reading gains are not moderated by the racial composition of the student body, and the growth in socioeconomic and racial gaps in reading gains is similar for grouped and non-grouped students when the sample is limited to students in integrated schools. Overall, and with the sampling limitations discussed in the text in mind, skill grouping appears to be used similarly – and to have similar effects on reading gains – in both segregated and integrated schools.