Ethnic differences in achievement growth: Longitudinal data analysis of math achievement in a hierarchical linear modeling framework

Author: Yun Xiang

Persistent link: http://hdl.handle.net/2345/676

This work is posted on eScholarship@BC, Boston College University Libraries.

Boston College Electronic Thesis or Dissertation, 2009

Copyright is held by the author, with all rights reserved, unless otherwise noted.
ETHNIC DIFFERENCES IN ACHIEVEMENT GROWTH:
LONGITUDINAL DATA ANALYSIS OF MATH ACHIEVEMENT IN A
HIERARCHICAL LINEAR MODELING FRAMEWORK

Dissertation
by
YUN XIANG

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

May, 2009
Ethnic differences in achievement growth:

Longitudinal data analysis of math achievement in a hierarchical linear modeling framework

By Yun Xiang

Henry Braun, Ph.D., Chair

Abstract

Given the call for greater understanding of racial inequality in student achievement in K-12 education, this study contributes a comprehensive, quantitative, longitudinal examination of the achievement gap phenomenon, with particular attention to the organization characteristics of schools and school districts. Employing data from a large number of districts in a single state, it examines the trends in achievement and the growth in achievement after the passage of NCLB. It focuses on mathematics performance from grade 6 to grade 8. Both a traditional descriptive approach and one employing Hierarchical Linear Models were applied and compared. The purpose was not to determine which methodology is superior but to provide complementary perspectives. The comparison between the two approaches revealed similar trends in achievement gaps, but the HLM approach offered a more nuanced description. Nonetheless the results suggest that it is useful to employ both approaches. As to the main question regarding ethnicity, it appears that even if student ethnicity is confounded with other indicators, such as initial score and socio-economic status, it is still an important predictor of both
achievement gaps and achievement growth gaps. Moreover, demographic profiles at the school and district levels were also associated with these gaps.
DEDICATION

To my daughter and best friend, Sophia Mao, for her love and support.
I would like to express my sincere gratitude for all the understanding, encouragement, support and patience I have received from my family, friends, and Committee through the dissertation process.

Completing this dissertation has been a humbling experience to me. First, I have been amazingly fortunate to have Dr. Henry Braun as my Committee Chair, who provided me with invaluable feedback and constructive criticisms. Dr. Braun devoted a lot of his time to making me think hard and pushing me every step of the way. My appreciation to his guidance and help is beyond description! Only now when I look back, I realize how much I have benefited from this challenging process. I will try to keep working towards the example he sets for me: think clearly, write concisely, and communicate effectively.

I would also like to thank the other two committee members, Dr. Damian Betebenner and Dr. Walt Haney, who have been extremely understanding and patient especially when I was away in my last year of the program. Dr. Betebenner, who was my advisor for my first two doctoral years, laid a foundation for me to understand the current methodological issues in the educational field. His support and encouragement always motivate me to work persistently. Dr. Haney, with his rich experience in testing, provided thoughtful feedback, raised good questions, and taught me to be a critical thinker. I have been blessed to work closely with such a great committee.

I am indebted to the Department Chair, Dr. Larry Ludlow. Since my first day in the program, Dr. Ludlow has provided me with warm encouragement and helped me greatly with financial support. I am also grateful to Dr. Ted Youn for his generous support and practical advice.

I am also thankful for Dr. Diana Pullin and Dr. Charles Baron who allowed me to pursue my curiosity beyond research methodologies. Their courses in educational law and constitution have been among the experiences I cherished most at Boston College.

I would also like to acknowledge Dr. Michael Seltzer and his student group at the University of California at Los Angeles (UCLA) for the lectures and discussions on related topics that helped me improve my knowledge in the area.

Many friends at Boston College have been with me through the highs and lows. Their care and support not only helped me through some difficult times, but also greatly enriched my life in Boston. I truly value their friendship and I deeply appreciate their belief in me. I can only name a few here: Yves Salomon-Fernandez, Rachel Kay, Dongning Bai, Qingwen Xu, Wei Tao, and Yang Wang.

The love of my family has been my backbone for all these years. My daughter, Sophia, made the dissertation work a lot of fun when we worked together at home or a coffee
shop. She kept my spirits up. My mother and father have trusted me all the time and had faith in my accomplishing every milestone in life. My sister, Ling, gave me the strongest support at one of my most needed times. I thank my husband, Yuanbing, for his patience and encouragement. I also thank my cousin, Jian and Can, for their hospitality and tremendous support since the first day I came to Boston.

Lastly, I appreciate the support from Dr. John Cronin, Dr. Steve Wise and Dr. Yeow Meng Thum in the Northwest Evaluation Association.
# TABLE OF CONTENTS

Acknowledgement ................................................................................................... i

List of Figures ..........................................................................................................v

List of Tables ......................................................................................................... vi

Chapter 1: Introduction ............................................................................................1

Chapter 2: Literature Review ...................................................................................7
  Achievement Gaps versus Ethnic Differences in Achievement Growth ................. 7
    Inequalities in Education .................................................................. 7
    Studies of Achievement Gaps .............................................. 10
    Approaches to Analyzing the Achievement Gap .................. 12
    Definition ............................................................................. 14
    Assumptions ................................................................. 15
  Schools and School Districts ............................................... 18
  Segregation ................................................................. 22
  Mobility ................................................................. 28
  Hierarchical Linear Modeling (HLM) ........................................ 30
    The Unit of Analysis ..................................................... 30
    Model Variance Components ......................................... 31
    Mixed Effects .......................................................... 32
    Standard Error .......................................................... 33
  Contributions of the Study ................................................. 34

Chapter 3: Research Design ...................................................................................38
  Data and Measurement .............................................................................. 38
    Data .................................................................................. 39
    Measurement .................................................................... 44
    Variables ......................................................................... 49
  Analytic Strategy ............................................................................... 55
    Descriptive Analyses ....................................................... 56
    Hierarchical Linear Modeling ........................................... 58
      Two-Level Model ......................................................... 58
      Three-Level Model ..................................................... 63
  Future Results and Discussion ................................................... 69

Chapter 4: Results ..................................................................................................71
  Descriptive Analyses ............................................................................... 72
    Achievement and Achievement Gaps .................................. 72
      Overall Level .............................................................. 72
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>District Level</td>
<td>85</td>
</tr>
<tr>
<td>School Level</td>
<td>90</td>
</tr>
<tr>
<td>Comparisons among Levels</td>
<td>94</td>
</tr>
<tr>
<td>Independent Variables</td>
<td>99</td>
</tr>
<tr>
<td>Hierarchical Linear Modeling</td>
<td>110</td>
</tr>
<tr>
<td>Preliminary Analyses</td>
<td>111</td>
</tr>
<tr>
<td>Two-Level Unconditional Model</td>
<td>111</td>
</tr>
<tr>
<td>Three-Level Unconditional Model (Level Three–School)</td>
<td>115</td>
</tr>
<tr>
<td>Three-Level Unconditional Model (Level Three–District)</td>
<td>119</td>
</tr>
<tr>
<td>The Hispanic-White Gap in Achievement Growth</td>
<td>121</td>
</tr>
<tr>
<td>Two-Level Conditional Model</td>
<td>121</td>
</tr>
<tr>
<td>Three-Level Conditional Model (Level Three–School)</td>
<td>134</td>
</tr>
<tr>
<td>Three-Level Conditional Model (Level Three–District)</td>
<td>137</td>
</tr>
<tr>
<td>Black-White Gap in Achievement Growth</td>
<td>142</td>
</tr>
<tr>
<td>Two-Level Conditional Model</td>
<td>142</td>
</tr>
<tr>
<td>Three-Level Conditional Model (Level Three–School)</td>
<td>148</td>
</tr>
<tr>
<td>Three-Level Conditional Model (Level Three–District)</td>
<td>151</td>
</tr>
<tr>
<td>Chapter 5: Discussion</td>
<td>156</td>
</tr>
<tr>
<td>Comparison between the Descriptive Analyses and the HLM Growth Model</td>
<td>157</td>
</tr>
<tr>
<td>The Concept of Ethnic Differences in Achievement Growth (Achievement Growth Gap)</td>
<td>158</td>
</tr>
<tr>
<td>The Relationship between Where Student Start and How Fast They Grow</td>
<td>162</td>
</tr>
<tr>
<td>Random Effects and Fixed Effects at Different Levels</td>
<td>164</td>
</tr>
<tr>
<td>Policy Implications</td>
<td>169</td>
</tr>
<tr>
<td>The Limitations of the Study and the Potential Directions for Future Studies</td>
<td>177</td>
</tr>
<tr>
<td>References</td>
<td>179</td>
</tr>
<tr>
<td>Appendix A</td>
<td>185</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure                                      Page

4.1.1  The Mean Scale Scores of the Three Ethnic Groups ........................................74
4.1.2  The Standard Deviations of the Three Ethnic Groups ......................................74
4.1.3  The Score Gaps between Ethnic Groups .............................................................76
4.1.4  The Effect Sizes of the Score Gaps between Ethnic Groups ..............................76
4.1.5  Mean Gain Scores across Performance Levels ...................................................83
4.1.6  Average Scale Scores of Each District for Different Ethnic Groups .................86
4.1.7  Average Gain Scores of Each District for Different Ethnic Groups over Two Years ...........................................................................................................87
4.1.8  Achievement Gaps between Black and White Students and between Hispanic
       Students at the District Level ..............................................................................88
4.1.9  Gain Score Gaps between Black and White Students and between Hispanic
       Students at the District Level ..............................................................................90
4.1.10 The Average H-W and B-W Scale Score Gaps in Every School .........................91
4.1.11 The Average H-W Scale Score and Gain Score Gaps for Different Schools ..........92
4.1.12 The Average B-W Scale Score and Gain Score Gaps for Different Schools ..........93
4.2.1  The Expected Growth Trajectory for Hispanic and White Students
       Based on Their Initial Status ..............................................................................133
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>The Data by Year and by Grade .................................................................39</td>
</tr>
<tr>
<td>3.2</td>
<td>Demographic Background of the Students in the Two Cohorts .................40</td>
</tr>
<tr>
<td>3.3</td>
<td>The Distribution of Racial Group in the Districts ....................................42</td>
</tr>
<tr>
<td>3.4</td>
<td>Student Mobility Index for Different Patterns of Moving ......................50</td>
</tr>
<tr>
<td>4.1.1</td>
<td>The Means and the Standard Deviations of the Overall Scale Scores ...........73</td>
</tr>
<tr>
<td>4.1.2</td>
<td>The Description of the Gain Scores of the Three Ethnic Groups ..........78</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Distribution of Performance Levels at the Starting Year by Ethnicity ....81</td>
</tr>
<tr>
<td>4.1.4</td>
<td>Models with the Dependent Variable Gain Score ..................................82</td>
</tr>
<tr>
<td>4.1.5</td>
<td>Comparing the Average Scale Scores at Three Levels – District, School and Overall ........................................................95</td>
</tr>
<tr>
<td>4.1.6</td>
<td>Comparing the Average Gain Scores at Three Levels – District, School and Overall ..........................................................96</td>
</tr>
<tr>
<td>4.1.7</td>
<td>Comparing the Scale Score Gaps at the School Level and the District Level ............................................................97</td>
</tr>
<tr>
<td>4.1.8</td>
<td>Comparing the Gain Score Gaps at the School Level and the District Level .............................................................98</td>
</tr>
<tr>
<td>4.1.9</td>
<td>The Index of Racial Diversity and the Distribution of Racial Groups in the Districts .........................................................101</td>
</tr>
<tr>
<td>4.1.10</td>
<td>Minority Percentage in Schools and Districts ......................................102</td>
</tr>
<tr>
<td>4.1.11</td>
<td>Percentage of the Students Who Stayed in Their Old Schools, Bureaucratically Moved, and Moved for Non-bureaucratic Reasons ..................102</td>
</tr>
<tr>
<td>4.1.12</td>
<td>The Other Independent Variable at the District Level ............................104</td>
</tr>
</tbody>
</table>
4.1.13 The Other Independent Variable at the School Level..............................104

4.1.14 Gender Differences in Student Achievement Scores...............................105

4.1.15 Descriptions of Different Groups of Movers...........................................106

4.1.16 Multiple Comparisons Regarding the Difference in the Average
    Scores...................................................................................................... 107

4.1.17 Correlations between the Achievement/Achievement Gaps and the
    Independent Variables at the School Level .............................................108

4.1.18 Correlations between the Achievement and Achievement Gaps and
    the Independent Variables at the District Level.................................109

4.2.1 Two-Level Unconditional Linear Model of Growth in Math
    Achievement ............................................................................................113

4.2.2 Three-Level Unconditional Model of Growth in Math Achievement
    across Schools..........................................................................................117

4.2.3 The Comparison between the Three Level Unconditional Models
    with Schools and Districts as the Third Level Units Respectively........120

4.2.4 Two-Level Conditional Model of Growth in Math Achievement with
    the Effect of Ethnicity ..............................................................................122

4.2.5 Variance Explained in Initial Status and Growth Rate as a Result of
    the Effect of Ethnicity..............................................................................124

4.2.6 Fixed Effects of the Two-Level Conditional Model with Initial
    Status Controlled for ................................................................................126

4.2.7 Fixed Effects of the Two-Level Model after Initial Status and the
    Interaction Effect between Ethnicity and Initial Status Being
    Controlled for (Compared with Model 2)................................................129

4.2.8 Fixed Effects of the Two-Level Conditional Model with Person-
    Level Predictors (Compared with Model 3)............................................130

4.2.9 Expected Growth Rates for Hispanic Students and White Students.......132

4.2.10 The Relationship between School Variables and Math Achievement
    (Three-Level Conditional Model)...............................................................135
4.2.11 Variance Decomposition from a Three-Level Analysis (Level 3 - School) ........................................................................................................................137

4.2.12 Relationship between District Variables and Math Achievement (Three-Level Conditional Model) ............................................................................139

4.2.13 Variance Decomposition from a Three-Level Analysis (Level 3 - District) .............................................................................................................140

4.2.14 Two-Level Conditional Model of Growth in Math Achievement with the Effect of Ethnicity .................................................................142

4.2.15 Fixed Effects of the Two-Level Conditional Model with Initial Status Controlled for .............................................................................................144

4.2.16 Fixed Effects of the Two-Level Conditional Model after Initial Status and the Interaction Effect between Ethnicity and Initial Status Being Controlled for .................................................................146

4.2.17 Fixed Effects of Two-Level Conditional Model with Person-Level Predictors .........................................................................................................147

4.2.18 Relationship between School Variables and Math Achievement (Three-Level Conditional Model) .............................................................148

4.2.19 Variance Decomposition from a Three-Level Analysis (Level 3 - District) .............................................................................................................152
Chapter 1: Introduction

The central part of this dissertation answers the query about how achievement gaps have changed through years after the No Child Left Behind act and to what extent the variation in ethnic differences in achievement growth can be accounted for by characteristics of the public school system.

Education quality and equity have been the two major themes of the various education reforms, policies, and laws over the last four decades (Coleman, 1966; Harris & Herrington, 2006). The most recent nation-wide effort was the No Child Left Behind (NCLB) Act of 2001 which requires every state to administer annual reading and mathematics tests to students in grade 3-8 based on their own curriculum and academic standards. The act aims to ensure that every student in public school system reaches the state-determined proficiency level by 2014. The reform emphasizes student achievement as measured by state-wide standardized assessments. By introducing school accountability, the federal government tries to play a role in the local k-12 education system. School districts and schools that fail to meet adequate yearly progress (AYP) towards state proficiency goals will be subject to different sanctions ranging from “improvement, corrective action, to being taken over or converted to a charter school” (No Child Left Behind Act, 2002).

In addition to the goal of raising the academic performance in general, the law pays special attention to children who are from disadvantaged populations. The adequate yearly progress (AYP) indicators reported by every school are calculated by breaking down the student population into different groups. These groups are defined by the categories of poverty, ethnicity, disability, and limited English proficiency. Thus, school
districts and schools are held accountable not only for overall academic performance, but also for academic performance of the specific groups. Closing the achievement gaps between different ethnic groups is gaining more and more attention since America has become more racially diverse and the achievement gap has remained stable. With a focus on racial equity, the initial research question of this dissertation is whether the achievement gap between minority and white students has been decreasing after the No Child Left Behind act was implemented. Note that the term the achievement gap(s) here specifically refers to the gap(s) between different ethnic groups instead of any other groups defined by the categories of poverty, disability, or limited English proficiency.

Although there is a wealth of research on patterns and trends in the studies of the achievement gap, most of the studies have been cross-sectional instead of longitudinal, largely because only recently has there been sufficient longitudinal data on student achievement. Student achievement in many studies, is quantified, and to some extent, simplified as a test score. The use of standardized test data as the means for assessing student progress and evaluating school effectiveness is quite controversial (Kane, & Staiger, 2002; Linn, 2000, 2003, 2004). While in cross-sectional studies the student achievement scores at one time point are compared, longitudinal studies with the test scores available at multiple time points generally adopt one of the three approaches: (1) Comparing the test scores of different cohorts in different years to understand the trend in the achievement gap based on the observations of different groups of students. A special case of the studies of different cohorts are the studies of successive cohorts which compare the test scores of the same grades but in different years, for example, comparing
the test scores of students in grade 5 in the year 2002 with the scores of those in grade 5 in the year 2003; (2) Tracking the same student population but different sample of students. One example of this repeated cross-sectional approach is the study of National Assessment of Educational Progress (NAEP). In its sampling design, fourth-grade students are randomly drawn from the target student population when four years later, eighth-grade students are randomly drawn from the same student population. (3) Tracking the test scores of the same cohort of students from one grade to another, for example, from grade 5 to 6 to 7 and beyond in three consecutive or nonconsecutive time points. According to Miller (2003), the most commonly used indicators for analyzing the achievement gap are percent achieving a performance standard (mostly percent achieving proficiency), mean scale scores, and effect sizes. The emphasis on school accountability in the NCLB act, however, has led more and more policy makers and other interest groups to view student achievement as a process of growth in a specific school or school district instead of simple decisions regarding passing(s) or failing(s) at one-time point or even multi-time points. Expanding the definition of the achievement gap becomes one major purpose of the dissertation. The new definition aims to help policymakers and other stakeholders better understand the issue of the achievement gaps so that they can formulate more effective policies and strategies. The distinction between the term achievement gap and ethnic difference in achievement growth will be elucidated. The results will be compared to see whether the distinctions have an impact on the conclusions with regard to racial inequity in student achievement as well as school accountability. The growth difference here refers to the average differences in growth
rates of the achievement scores between minority and white students, while considering the relationship between growth rates and growth intercepts. Hence, the concept of growth in this study includes the three elements: (1) where students start (growth intercept); (2) how they progress (growth rate); (3) the relationship between growth rates and intercepts.

Although the country is becoming more racially diverse, the efforts spearheaded by the civil rights movement decades ago, including school desegregation and affirmative action, are being challenged (Bowman, 2001; Lee, 2003; Orfield, 2005, chap 1). Under the circumstances, the change of racial distribution in school districts and schools must be more fully understood. Moreover, the recent trends in school resegregation, suggest that an examination of trends in the achievement gap(s) is timely. Therefore, before describing the analyses intended to answer the general research question, the dissertation will first examine the trends in racial distribution in schools and districts, providing a frame for the study in the context of the recent literature on school segregation. The study then uses a particular segregation index and the measure of racial composition as the indicators of racial diversity in schools and districts. The measures of racial diversity combined with other characteristics of schools and school districts will be included in the model to explain the variance in achievement growth, as well as the variance in ethnic differences in achievement growth.

Given the calls for an expanded understanding of racial inequality in student achievement in our k-12 public school system, particularly for studies that examine
organizational effects\(^1\) of schools and school districts, the current research seeks to contribute a more comprehensive, quantitative, longitudinal understanding of the phenomenon of racial inequality in education achievement. The purpose of this dissertation is to examine the trends in the achievement gap and the achievement growth while investigating organizational effects of the schools and the school districts. The study will focus on student mathematics achievement growth from grade 6 to grade 8.

The relevant research questions include:

- Have the achievement gaps in different grades changed through the three academic years from 2002 to 2005? (Descriptive data analysis)
- Have the achievement scores of minority students (Black and Hispanic students) grown faster on average than the scores of White students? (Two-level Hierarchical Linear Model analysis)
- Do the ethnic differences in achievement growth vary across schools or districts, and if so, which school or district factors are associated with such variation? (Three-level Hierarchical Linear Model analysis)

In addition to conducting analyses for each research question, the results from the descriptive analyses of the first research question regarding the achievement gap will be compared with the results from the Hierarchical Linear Modeling (HLM) analyses of the second research question regarding the achievement growth differences. The value of the dissertation, to a great extent, rests on the comparison between the two analyses.

\(^1\) Note that the term *organizational effect* does not imply causal effect. It only refers to how much variation of achievement growth can be accounted for by the random variation across schools or districts.
The dissertation is organized as follows: The next sections will provide an extended review of the literature. The data and methods will then be described after introducing the particular contributions of this study. The results section will contain the basic descriptive results and the multi-level analyses of Hierarchical Linear Model. The dissertation concludes with a discussion of findings, implications, and the limitations of the study.
Chapter 2: Literature Review

“Hence it is the business of education in a democratic social group to struggle against this isolation of social groups and classes in order that the various interests may reinforce and play into one another” (Dewey, 1916, p.292).

The literature review addresses the meaningfulness of this study by examining important aspects of the dissertation. First, a number of studies are introduced to document the trends in the achievement gaps. Traditional approaches to analyzing these gaps are critiqued. The term *ethnic difference in achievement growth* is then defined and compared with the term *achievement gap*. The fact that students are nested within schools and districts introduces the question of the organizational effects of schools and school districts. Moreover, to explain what factors can account for the variation in student achievement growth and ethnic differences in growth, two important predictors—indexes of school segregation and mobility are discussed. Lastly, the advantages of the methodology of Hierarchical Linear Modeling (HLM) are highlighted.

**Achievement Gaps versus Ethnic Differences in Achievement Growth**

*Inequalities in Education*

Inequalities in education across racial and ethnic groups have long troubled those who “see education as a way of reducing social disparities by compensating for past injustices and countering present social inequalities” (Hallinan, 2001). Before African American students had freedom to attend the same schools with White students, educational research and policy had focused on educational opportunities instead of educational achievement or attainment. In the landmark case *Brown v. Board of*
of Topeka (1954), the Supreme Court outlawed unequal educational opportunity by unanimously agreeing that segregation in public school is unconstitutional. The ruling overturned the doctrine of “separate but equal” in Plessy v Ferguson (1896) fifty years before, where the majority of the judges in the Supreme Court held that students had the equal educational opportunity, even if Black students were separated from White students by going to different schools. Although the landmark Brown case paved the way for the desegregation of public schools, the decision actually encountered a great deal of resistance, especially from southern states. The most famous example was that in 1957 in a formerly all-white high school in Little Rock, Arkansas, nine black students were blocked from entering the school on the orders of the Governor. President Eisenhower had to send federal troops to intervene on behalf of the students. Since the 1960’s, the civil rights movement has made great progress. The Civil Rights Act of 1964, prohibited discrimination of all kinds based on race, color, religion, or national origin. The act outlawed segregation in public facilities by providing the federal government with the power to enforce desegregation. Children in public school, no matter what race/ethnicity they are, began to gain equal access to educational facilities by going to the same school districts, schools and classes. The progress has gradually drawn attention to a significant gap between the achievements of non-Hispanic White (European American) and minority students, especially Black (African American) students. The consistent gaps themselves become a disturbance to “the effective functioning of a democratic, technological, diverse society such as the American one” (Miller, 1999).
Studies also found that the achievement gaps are related to many social problems, such as unequal job opportunity, lower family social-economic status, etc (Kirsch, Braun, Yamamoto, & Sum, 2007). Kirsch, et al. (2007) also claimed in their report that economic opportunities would not improve without more educational reforms and efforts be made. The three forces which they believed are consequential to the nation’s future include substantial disparities in skill levels (reading and math), seismic economic changes (widening wage gaps), and sweeping demographic shifts (less education, lower skills). In the dissertation, two out of the three forces will be discussed by analyzing what the achievement gaps in mathematics are and how the gaps are correlated with the current demographic characteristics of students.

Believing that education should play a positive role in the American process of democracy, Chief Justice Earl Warren addressed the importance of education in Brown (1954, p493):

Today, education is perhaps the most important function of state and local governments. Compulsory school attendance laws and the great expenditures for education both demonstrate our recognition of the importance of education to our democratic society. …It is the very foundation of good citizenship. Today it is a principal instrument in awakening a child to cultural values, in preparing him for later professional training, and in helping him to adjust normally to his environment. In these days, it is doubtful that any child may reasonably be expected to succeed in life if he is denied the opportunity of an education.

To reduce inequalities in education and close the achievement gap, the federal and state governments have made efforts such as providing funding for educating “educationally disadvantaged” children, raising academic standards, and making state and local policies more coherent. The efforts made at the federal level, such as re-
authorization of ESEA (Elementary and Secondary Education Act) and the No Child Left Behind Act, work together with various initiatives undertaken by different states. The most recent nation-wide provisions of the No Child Left Behind Act (NCLB) mandated the allocation of funds to ensure that every child, especially those disadvantaged children, meets a certain standard. Furthermore, additional resources are provided to the districts with large proportions of students from poorer families, many of whom are minority.

Studies of Achievement Gaps

For decades, educational researchers and policy makers discussed and crafted proposals to close the achievement gap between minority and non-minority students (Bainbridge & Lasley, 2002; Coleman et al., 1966; Harris & Herrington, 2006; Henderson, 1975; Jencks & Phillips, 1998). However, the problem remains large and growing despite substantial interest and effort from policy makers and educators. A review of the research indicated that the achievement gap between white and minority students narrowed significantly from the 1960s to the mid-1990s (Hedges & Nowell, 1998; Ipka, 2003; Jencks & Phillips, 1998; Lee, 2002). However, both Hedges and Nowell (1998) and Jencks and Phillips (1998) reported that the rate of decrease had slowed since the 1970’s when the average Black students still scored lower than White students on most standardized tests.

The decrease in the achievement gap in both reading and mathematics is indicated in the long-term trend component of the National Assessment of Educational Progress (NAEP), a survey begun in the 1970s to assess trends in students in reading, mathematics, science, and writing in the 4th, 8th, and 12th grades. The NAEP data revealed
that the achievement gap between Blacks and Whites during the 1970s and 1980s narrowed; but the gap began to widen in the 1990s. The most recent analyses of data from the NAEP (National Center for Education Statistics, NCES, 2005) in Mathematics indicated that the achievement gaps between Black and White and between Hispanic and White students in mathematics is present in elementary school and continues through high school. The results also showed that the White – Black score gap at both grades 4 and 8 was narrower in 2005 than in previous assessments. However, the gaps in 2005 were not statistically significantly narrower than the gaps in 1990 (National Center for Education Statistics, 2005).

While NAEP shows that the achievement gap may have reached a plateau during the 1990’s, the Early Childhood Longitudinal Study (of kindergarten cohort) draws different conclusions. This longitudinal study includes a nationally representative survey of more than 20,000 children who entered kindergarten in 1998. Overall, non-Hispanic White students in kindergarten and 1st grade were found to score higher than their Black and Hispanic peers. Moreover, the achievement gap in math between Black and White students was found to be smaller when compared with the results of previous studies. These seemingly contradictory findings, based on different samples of students, sparked researchers’ interest in investigating the current trend in the achievement gap, especially after passage of the NCLB act.

Believing the achievement gap still “remains a defining mark of racial inequality in public education today” (Hallinan, 2001), this study will focus on the achievement gap in mathematics to examine the link between student achievement and race-ethnicity. The
term achievement gap is typically used to refer to the gap between white students and Black students. The bulk of the research on the achievement gap has focused on the patterns of the Black-White achievement gap. The underlying assumptions of these studies is that the patterns and trend in the achievement gap as well as the factors explaining the achievement gap are sufficiently similar for all minority groups (Bowman, 2001; Carpenter, Ramirez & Severn, 2006). As the Latino group recently became the largest minority group, using the term achievement gap to refer primarily the difference between Black students and White students is no longer appropriate. A narrow definition of achievement gap is be insufficient for today’s school population.

**Approaches to Analyzing the Achievement Gap**

Before we further investigate the achievement gap, the definition of the achievement gap should first be examined. In previous studies of the achievement gap, the attention typically was given to the *static* differences between white and minority students instead of the differences of the achievement *growth*. The static difference here refers to the comparison at one time point of achievement scores at an aggregated data level (i.e., the cumulative results at the school, district, or the state level). Two-wave studies are only marginally better by computing and comparing the *gain scores* by year and/or by grade. They are inadequate for studying change since the information about the shape of individual’s growth trajectory is missing (Singer & Willett, 1996). Some researchers began to focus on the trajectory of the achievement gap across multiple time points (Orfield, 2005, chap 6). Braun, Wang, Jenkins, & Weinbaum (2006), however, realized that this approach might lose the information on how the achievement of both
White and minority (Black) students varies over time. In particular, they considered both absolute gains by minority (Black) students and the progress in closing the achievement gap. The progress, though based on the data of at least three consecutive or non-consecutive time points, could easily be confounded with student background characteristics. Such inter-group comparisons allow the question to be raised about student comparability over time since they are based on the assumption that students enrolled in different years would be very similar from cohort to cohort.

Both cross-sectional studies and longitudinal studies based on different groups of students confound cohort effects with real difference in student achievement, and thus are subject to selection bias. Cohort effects take place when observed differences in student achievement may be largely due to the differences in background characteristics and previous achievement instead of school performance or program effect.

In sum, the literature describes four approaches to investigate the achievement gaps: (1) static difference—compare one-year test scores of different ethnic groups; (2) gain score—compare the gain scores across two time points; (3) the progress—to compare the test scores of at least three consecutive or nonconsecutive years based on different cohorts of students; (4) the growth—to compare the test scores of at least three consecutive or nonconsecutive years based on the same cohort of students. Conceptually, the fourth approach, the growth modeling of student achievement better represents the time-dependent process of academic learning, effectively excludes the cohort effects, and provides a degree of control over student background characteristics (Seltzer, Choi & Thum, 2003; Willett, 1988). It is the approach which will be applied in the dissertation.
**Definition**

The analyses presented in this study are based on the growth of one cohort of students. Therefore, the term achievement gap comes from a longitudinal instead of cross-sectional perspective. The study aims to investigate how the achievement scores of different ethnic groups grow over years. The growth rate and intercept instead of the static difference in the achievement scores becomes the focus of the dissertation. *Hence, ethnic difference in achievement growth refer to the difference in achievement growth rates and intercepts between white and minority students based on observations of the same cohort of students at multiple time points.* This definition is an effort to help policymakers and others better understand the issues of the achievement gaps and racial inequality. The phrase “achievement growth” here does not necessarily mean that the outcomes must “grow” or increase over time. The term is applied without considering the specific direction of change. Some studies use the terms initial status and rate of change to replace the term growth intercept and growth rates (Seltzer, Choi, & Thum, 2003). In the dissertation, both terms are used interchangeably.

The necessity of taking into consideration the location of the score was also addressed by Rock & Pollack (2002) in their study of early literacy education. The study showed that the traditional approaches to measuring change, such as raw gain score or gains from ANCOVA with the pretest as a covariate, may yield misleading results. For example, students starting at a higher score may not increase their scores as fast as those starting at a lower score when these scores are reaching the ceiling. Hence, by including the location of the scores, they introduced the notion of the percent of maximum possible
gain to minimize the impact of ceiling effects. The results have an important implication: when measuring change, it is important to take into account “where on the vertical scale the gain was taking place as well as the amount of gain”.

Generally, if the waves of data only contain three or four time points, the growth is assumed to be linear over time since with such few time points the shape of the projection is not very revealing. The individual growth model contains two growth parameters - an intercept and a slope - representing an initial value and a rate of change (Singer & Willett, 1996, 2003; Raudenbush & Bryk, 2002, chap 6). Hence, the variance in intercepts and slopes can both be revealing. Seltzer, Choi & Thum (2003) further pointed out that in addition to investigating these two growth parameters, it is also important to consider the relationship between students’ initial status (where students start) and the rate of change (how rapidly they progress).

**Assumptions**

There are several important assumptions in the growth modeling approach in the dissertation:

First, the notion of achievement gap is specified to be the score-gap based on the standardized tests in a state assessment system. The assumption is made that the state assessment system assesses student learning with adequate validity. Issues such as test quality, cheating, score inflation, etc, are not addressed here. The focus is describing student progress and comparing achievement growth. However, these issues are good reminders that test scores do not completely capture student learning, and thus inferences based on test scores should be made carefully and cautiously.
Second, linking becomes a critical issue here since the assumption is made that the scale scores across years and across grades have the same meaning and reflect the real progress of student learning. Item Response Theory (IRT) models are used to estimate item parameters and adjust them to a common scale. According to Vale (1986), linking consists of the two elements - an anchoring design and a transformation method. There are a number of methods which transform item parameter estimates in one metric to another. The Stocking and Lord (1983) procedure is applied in the state assessment system to link the parameter of the items onto a common scale. The method requires that some common items appear in different tests and are administered to different groups of examinees. All common items, by being administered to each examinee, are used to produce an estimated true score which is computed as the sum of the ICC (Item Characteristics Curve) probabilities. By addressing the standard error of estimate, this approach has been widely used in various studies. However, Clemans (1993) made a comparison about the effectiveness of IRT and Thurstonian scaling procedure and found the latter more accurate. He correctly pointed out some limitations or assumptions underlying the IRT model. For example, the assumption of unidimensionality was made (i.e. all items measure a single trait). This assumption was also mentioned by Stocking and Lord (1983), who advocated the IRT approach. Clemans’ suspicion of the inferiority of the IRT approach, however, was largely based on the observation he found—“the variance of the equal-interval scores decreases as a function of grade level”. In fact, there exist some alternative ways to explain this phenomenon. For example, at higher grades, it may be easier for students at the lower end, compared with those at the upper end, to
improve their test scores (ceiling effect). Clemans was not the only one who found differences in variance trends for different scaling methods. A frequently discussed example was provided by Yen (1986), where Thurston scaling was compared with the IRT 3-parameter scaling when both were applied to the California Achievement Test. The differences found between the two approaches stimulated a sharp debate in the psychological field over what scaling method was more accurate. However, no specific conclusion was made from the debate (Hoover, 1988 & Yen, 1988). All these studies remind researchers and policy makers that scaling procedures should be carefully evaluated especially when the items were too easy or too hard to reflect the entire range of ability. Moreover, the results based on these procedures should be cautiously used.

Third, Asian students have generally performed better in mathematics. Thus, the gaps in this subject mainly remain between non-Asian minorities and the white majority (Miller, 1999). Therefore, the dissertation will only focus on the achievement gaps between non-Hispanic White students and African American (Black) students as well as Hispanic students. The minorities in this study, therefore, only refer to non-Asian minority students. Moreover, to simplify the terms, we sometimes refer to non-Hispanic White as White and African American as Black. Hispanic refers to those students who “self-identify or share the cultural attributes with one or more Latin American societies” (Miller, 1999). An assumption exists here that students are clearly labeled and the categories of race/ethnicity are mutually exclusive, which is usually not the case. The complexity of the categories makes the picture of the achievement gap even more complicated. For example, it is controversial for students with multi-ethnicity background
to define their category of ethnicity. Therefore, the conclusions based on the simple
categories in this study, although bringing the convenience to the analyses, need careful
and cautious interpretations.

**Schools and School Districts**

While education achievement and attainment draws more and more attention, a
part of the current discussion was found to focus on the failure of African American
students instead of systematic unequal educational opportunity (Love, 2004). James
Coleman (1994) claimed that the focus of the research on race and schooling must be the
social system rather than the individual. Hallinan (2001) further pointed out that based on
the existing theories of racial inequality, researchers should “take into account
individuals, schools, and communities and how they interact as a dynamic social system
to examine racial inequality” The most important but controversial aspect in this dynamic
social system is the school. The federal NCLB legislation requires schools and school
districts be held accountable for the performance of all children including disadvantaged
students and the minority students. The increasing demand to hold schools accountable
for their effects on student outcomes has drawn attention on the school effect. The NCLB
assumes that there is school effect and that it can be improved by threat of sanctions.
However, the term school effect can be misleading when researchers tend to make causal
inference on school characteristics (Goldstein, 1991). The absence of experimental
design, specifically speaking, without random assignment and treatment/control groups,
no causal interpretation should be made. In observational studies, the inference of school
effect should then be very carefully drawn especially when schools are held accountable for student performance or when they are ranked for rewards or sanctions.

Although more and more longitudinal data are available after passage of the NCLB act, states are not taking full advantage of the annual testing data. In order to measure student progress and school effectiveness, states are only required to utilize a status-type measure (i.e. the percentage of proficiency or above on the state test). Under NCLB, proficiency above a certain level is widely used in the state-level reporting system. School improvement then becomes the rate at which this percentage increases. However, researchers have pointed out that the core element of this accountability system - the cut-score used to define student proficiency level is very misleading and easy to manipulate for different purposes (Lion 2000, 2004).

Among various approaches to investigating school effects, the Value Added Model (VAM) has drawn a lot of attention since a number of states and some districts have adopted accountability systems based on it. The central theme of this approach is that schools should and can be held accountable for student learning (Sanders, Saxon, & Horn, 1997). In this system, students are tracked when they move from school to school in order to ensure that they are exposed to the “treatment” - the school. A school’s value is added when the rate of student learning in a school is increasing. The term value added strongly implies causal relationship between school effectiveness and student learning: the school adds value to what the child has known. Therefore it is often regarded as superior to other existing approaches. However, the causal reference cannot be valid since (1) test scores do not necessarily reflect the true difference in learning; and (2)
without a true experimental design, there are many other confounders which can explain the change of learning. Raundenbush (2004) said in his report about the value-added design: “if snapshots of average proficiency cannot reveal school quality, then changes in those snapshots cannot reveal school improvement.”

**In this study, the school effect is more of an estimate of the residuals at different levels; it depends on which variables are put in the model and thus may contain large standard errors and bias** (Aitkin, Anderson, & Hinde, 1981; Goldstein, 1991; Raudenbush, & Williams, 1995). The term school effect here is to help understand whether or not school and school district systems provide a positively stimulating environment for high-level academic achievement for minority students. Further, in order to discover which factors are associated with success or failure of schools, researchers have conducted various studies to identify the school effect (Lee, 1986; Raudenbush & Williams, 1995).

An array of variables, which are believed to be related to student achievement in mathematics, have been examined. The analysis conducted by Raudenbush & Bryk (1986) showed that the relationship between socio-economic status (SES) and mathematics achievement varies substantially across U.S. high schools. The Coleman Report (1966) claimed that in addition to the student’s own social background, there is a school factor which is called the *social composition* of the student body. The measures of social composition in the previous studies include the percentage of minority students and the percentage of students on free or reduced lunch (Lee, 1986; Orfield, 2005, chap 6). In the study on the effects of school organization and school size on changes in student
achievement during high school, Lee (2002) used two measures of social composition -
the mean SES of students in the school and high-minority schools (schools with 40
percent or more back and Hispanic students). They were found to be significant
predictors of the student achievement. In summary, the existing research suggests that
social composition significantly impacts student achievement (Lee, 2002; Muller, Stage,
& Kinzie, 2001; Orfield, 2005, chap 6; Raundenbush & Bryk, 1986).

In this study, eligibility for the federally assisted meal program - the National
School Lunch Program, serves as a measurement of social composition of the schools and
districts. To be eligible for free lunch, a student must be from a household with an
income at or below 185 percent of the federal poverty guideline; to be eligible for
reduced-price lunch, a student must be from a household with an income at or below 130
percent of the federal poverty guideline. Other variables, such as the school/district size,
pupil-teacher-ratio, percentage of minority students, segregation index, mobility level,
will also be included to give a comprehensive view of the social composition of schools
or districts.

Two explanations have been offered for the effects of social composition on
student achievement (Orfield, 2005, chap 6). The first explanation is peer influence. The
second is indirect effects on organizational and structural features of schools. For
example, minority students are more likely to attend large, high-poverty schools with
worse educational facility and fewer qualified teachers. The second explanation is
popular while investigating the organizational effects of schools or districts. Many
researchers attribute the inequality of education achievement to the difference in the
amount of resources that minority and non-minority students obtain. They tend to think that the quality of American public schools was shaped by the amount of wealth in every school district since American schools were funded primarily by local property taxes (Condron & Roscigno, 2003; Orfield, 2005, chap 1). For example, wealthier districts can tax more than poor districts and have more to spend on education. The differences in the amount of resources between districts make it necessary in this study to treat school district as a data level in the multi-level model.

Schools and school districts will be put in separate but parallel models in the study of multi-level modeling because: 1) The distinction between schools and districts is important and they have different policy implications; 2) We are interested in investigating the separate organizational effects of schools and districts to answer the question— which school or district characteristics are associated with the growth of student achievements as well as with the ethnic differences in achievement growth?

Segregation

In the past several decades, the U.S. population has become more racially ethnically diverse. In 1970, the U.S. population was 87 percent non-Hispanic white; by the mid-1990s, the population was 71 percent white (Littman, 1998). According to the U.S. Census Bureau, minority groups now constitute one third of the U.S. population. They further indicated there will be substantial change in the structure of population over the next few decades. For example, Hispanics and Asians will triple their current number of population by 2050, and by then non-Hispanic Whites may drop to half of total
population. The growth of metropolitan areas, particularly the suburbs, has been due largely to the increase in minority populations. The racial changes are perhaps most evident in the public schools because the school-age population is substantially less white than the total population: In 1972, only 22 percent of public school students were minority group members. In 2005, 42 percent of public school students were considered to be minority (National Center for Education Statistics, 2007). It is important to examine whether - and to what extent - the growth of minority population leads to integrated or segregated school environments for all students. The study conducted by Reardon and Yun (2001) found that after controlling the overall growth in enrollments, the increase in minority enrollment is associated with the increase in suburban segregation from 1987 to 1995. This is consistent with the statistics presented in Orfield’s study (2005) where he found that the percentage of African American students attending majority white schools has steadily decreased since 1986. This indicates that minority students are increasingly concentrated in highly segregated schools and school systems. He also pointed out that issues of segregation and equal opportunity, in the 1960s and 1970s, focused almost exclusively on the African American group; now attention must be paid to Hispanic and Asian populations as well.

As members of an increasingly diverse society, all children must be given the opportunities to interact across racial and ethnic lines. Isolation only contributes to the achievement gaps between White and minority students (Ikpa, 2003). As to the impact and consequence of isolation and segregation, the words of Chief Justice Earl Warren should be considered:
Segregation of white and colored children in public schools has a detrimental effect upon the colored children. The impact is greater when it has the sanction of the law; for the policy of separating the races is usually interpreted as denoting the inferiority of the Negro group. A sense of inferiority affects the motivation of a child to learn. Segregation with the sanction of the law, therefore, has a tendency to retard the education and mental development of Negro children and to deprive them of some of the benefits they would receive in a racially integrated school system. (Brown v. Board of Education of Topeka, Kansas, 347 US 483, 1954.)

Although the landmark Brown v. Board of Education of Topeka decision (1954) outlawed the separate educational facilities, the decision and other policies designed to assist minority group members in obtaining equal educational opportunities continue to be challenged by school districts throughout the nation. The challenge has led to the existence of many newly-resegregated schools of today. As Orfield (1999) pointed out, since the mid-1970's a lot of the American public schools have reversed the trend in desegregation and quietly resegregated. Meanwhile, studies since the 1970’s have found that there has been a strong link between the student achievement gap and the degree of segregation of the schools and school districts. The achievement gap between Black and White students decreased after the Brown case (1954) but stopped decreasing at the end of 1980’s. By then, some advocated the doctrine of “separate but equal” again that the students of different races gained equal educational opportunities even they were separate in different schools. Studies further found that children attending racially integrated schools performed better than those in segregated schools (Hubert, 1999; Ipka, 2003). The link between segregation and the achievement gap showed that the impact of school segregation on the academic achievement should be investigated more extensively. Hence, among various characteristics of schools and school districts, special attention
will be given to the investigations of the relationship between the degree of segregation and the student achievement growth differences.

Orfield (2005, chap 1) provided an explanation for the resegregation by stating that “desegregation policies have been largely abandoned because of declining support for desegregation from the executive and judicial branches of the federal government and the growing concentration of minorities in urban school districts that made meaningful desegregation nearly impossible.” An example given by him was the Swann v. Charlotte-Mecklenburg Board of Education (1971) case where the Supreme Court held that district courts have authority in “formulating remedies” in desegregation cases. Orfield then concluded that “resegregation will increase due to the nationwide trend in federal courts ending desegregation efforts in the first decade of twenty-first century.” The most recent case Parents v. Seattle School District (2007) confirmed his conclusion when the court ruled that race cannot be a factor in the assignment of children to public schools. Under such circumstances, this study aims to address the questions such as how serious segregation is now in suburban school districts and how suburban school segregation is associated with student achievement growth.

To measure how evenly students are distributed among districts by race, we select a measure of segregation. The choice of a segregation index directly influences the interpretation of the relationship between segregation and student achievement growth. There is a considerable literature on the merits and flaws of a variety of measures including different indices (Cortese, Falk & Cohen, 1976; Coltfelter, 1999; Duncan & Duncan, 1955; Reardon & Firebaugh, 2000; Reardon & Yun, 2001; Rumberger &
Williams, 1992; Zoloth, 1976). The concept of segregation in the literature is complex and somewhat “fuzzy.” Duncan & Duncan (1955) pointed out that on one hand, segregation indices can be an “informative aid in civil rights enforcement” when they are used to identify the problems and the progress as well as to direct the expenditure programs; On the other hand, researchers should be concerned about “arbitrary operationalization of the concept of segregation”. Hence, Duncan & Duncan stated that no single index may be sufficient because of the complexity of the notion of segregation. This study aims to identify appropriate measures of segregation to investigate the relationship between segregation and student achievement growth differences. Therefore, measures of segregation will be used as both a descriptive device and an important indicator of the district characteristic in the models applied.

As the country has become more racially diverse, the “two-group measures of segregation” become increasingly inadequate for describing complex patterns of racial segregation (Reardon & Firebaugh, 2000). In Reardon and Firebaugh’s study (2000), six measures of “multi-group segregation” were compared and further evaluated against a set of desirable prosperities of segregation indices. Among the indices, the Dissimilarity Index (D) is the most widely used one in the literature on school and residential segregation. Its interpretation is straightforward for education policy since it equals the proportion of minority (or nonminority) students who would have to be transferred in order to achieve the same racial composition in all schools. The index D is based on the absolute deviation of the racial composition of a school from that of the school district. It has a range of 0 to 1 and indicates the proportion of students that would have to change
schools to achieve an even distribution of students across all schools in a given school
district (Zoloth, 1976). The formula for D is

\[ D = \frac{\sum_{i} T_{i} |p_{i} - p|}{2Tp(1-p)} \]

where \( T_{i} \) and \( p_{i} \) are, respectively, the total enrollment and proportion minority of
the ith school, and where \( T \) and \( p \), respectively, are the total enrollment and proportion
minority of the district.

Although it is the easiest to compute the index D since it only requires the
information of the numbers of minority students and nonminority students in the two
groups of schools, Cortese, Falk & Cohen (1976) demonstrated that employing D would
lose the information on the particular proportion of nonwhites in an area. Thus, index D
is not appropriate to use when the number of the population of other minority groups,
instead of African Americans only, become significant. Therefore, in this study index D
is not chosen since the number of schools differs a lot among the districts and there is
considerable variation in the nonwhite proportion. Many researchers concluded that
among the segregation indices discussed in the literature, the one derived from the
information theory index H was the most “conceptually and mathematically satisfactory”
index (Theil & Finizza, 1971; Reardon & Firebaugh, 2000; Zoloth, 1976). The formula
and other details of the H index will be presented in the section on research design.
Mobility

Special attention is also given to mobility due to the research interest in investigating (1) whether the suburban schools in this study have a stable of cohort of students so that their growth can be tracked over time; (2) the unique contribution or prediction of the variable ethnicity after adjusting the effect of school mobility as well as other school characteristics; and (3) the organizational effects of schools. There is an implied assumption that students will attend a specific school or district consistently enough that the school can make a difference in their achievement. In addition to investigating whether students attend the same school, school mobility needs to be examined.

A variety of studies have been conducted to investigate the impact of the level of mobility on students, and more broadly on the classrooms and schools they attend. The negative associations between student mobility have been found for student achievement, grade repetition, and high school completion (Benson, Haycraft, Stayaert, & Weigel, 1979; Dunn, Kadane & Garrow, 2003; Kerbow, 1996; Haveman, Wolfe, & Spaulding, 1991). Controversies, however, exist. Heinlein & Shinn (2000) found that in their study of sixth-graders’ achievement, mobility was not related to subsequent achievement when prior achievement (third-grade achievement) was controlled.

As to the mobility effects on continuing students as well as on the schools which have mobile students, Heywood, Thomas, and White (1997) found “no evidence that mobility of classmates lowers achievement of stable students.” Other researchers disagreed by stating that the mobility effects on classroom instruction affected the
progression of subjects beyond the particular class and even across grades (Kerbow, 1996; Kirkpatrick & Lash, 1990). They further explained that a school that had mobile students would confront a dual task: maintaining the original pace for its continuing students and incorporating the mobile students without sacrificing the learning of the other children. Consequently, the mobility effect on continuing students will be considered in this study by including school mobility index in the model.

The existing studies of the effect of student mobility on achievement tend to indicate that a decline in achievement is generally associated with mobility (Benson, Haycraft, Stayaert, & Weigel, 1979; Dunn, Kadane & Garrow, 2003; Kerbow, 1996). However, very few studies were found to investigate the effects of several school transfers since longitudinal data on student achievement were lacking. In the British National Child Development Study, Blane (1985) demonstrated that children who had attended three or more schools, as compared to children who attended one or two schools, performed more poorly on measures of math ability after controlling their socio-economic level. In our three-year longitudinal data analysis, we will investigate how many times students changed schools and districts. Our hypothesis is that the effects of mobility may accumulate over time, and thus frequent school changes become additional impediments to student achievement growth. Accordingly, mobile students will be compared with continuing students; and the students who moved once will be compared with those who moved twice in three academic years.
Hierarchical Linear Modeling (HLM)

Hierarchical Linear Modeling (HLM) is used to address our hierarchically structured data (often called multi-level data). The term Hierarchical Linear Modeling first appeared in the paper by Strenio, Weisberg, & Bryk (1983). Other studies involving multi-level data include multilevel mixed linear models (Goldstein, 1986; Mason, Wong, & Entwistle, 1983), random coefficient models (DeLeeuw & Kreft, 1986), and hierarchical linear models (Bryk & Raudenbush, 1987). The term HLM is applied in the dissertation to capture the important data structure and addresses the research interest in the studies of growth and organizational effects. Based on the literature review, the advantages of applying HLM include:

The Unit of Analysis

The unit of analysis here refers to the research subjects. The most frequent research subjects in education studies are students who usually are assumed to be independent observations. However, students are dependent subjects since they are nested within various organizational units such as classrooms, schools, or school districts (Raudenbush, 1988). The dependence is due to the fact that when students are exposed to the same teacher, curriculum, and school climate, they tend to be more or less alike. The similarity of the units in a same organization makes it necessary to investigate this dependence. Moreover, even in experimental studies, educational researchers have found themselves facing a dilemma - It is not realistic to assign individual students, instead of the whole class or even the whole school, into a treatment group or a control group. This special feature of educational research, to be more specific, the dependence of the main
subjects (students) in various studies, provides a great rationale for the HLM analysis to address the problem of the unit of analysis. In our study, the first-level unit - the test scores are dependent since a student’s previous performance can be used to predict how s/he performs the next year. The same is true for the unit of second level - student. Students are more alike if they are in the same school or the same school district. HLM models can then be applied to investigate the within-unit and between-unit variance (Raudenbush, 1988).

**Model Variance Components**

In order to estimate school effects or academic program effects, very often the means of the outcome measures are compared. The comparison of the mean outcomes, to a great extent, neglected the real differences among different organizational units. The situation that students are nested in schools and districts has been simplified, and thus comparison based on it can lead to an invalid conclusion (Raudenbush & Bryk, 1987). Moreover, there is always an implied assumption of the homogeneity of variance in these studies, which is often not the case.

HLM is an effort to take into account the differences among organizational units (i.e. person, classroom, school, etc) by connecting residuals of variance from a lower level with the random “effect” of the organization units at a higher level. Including heterogeneity of variances helps explain the variation of the outcome variables while most of studies simply focus on estimating the mean differences (Raudenbush and Bryk, 1987). Hence, the variance instead of the mean becomes the outcome and focus of a study. The dispersion of the variability across organizational units is taken into the real
consideration. This is why the basic HLM model is also called *variance components analysis*. In this dissertation, the residual variance of student math achievement is modeled as a random variable at the different levels of the data structure. Besides estimating the variance of the parameters, HLM models also explain the residuals in student achievement as a function of individual or school characteristics. Educational research mostly is based on observational studies. Thus, heterogeneity of variance may arise either from program/treatment effects or from selection bias (Raudenbush and Bryk, 1987).

**Mixed Effects**

Multilevel mixed linear models have been applied to deal with hierarchically-structured data (Goldstein, 1986, 1987; Mason, Wong, & Entwistle, 1983). The term *multilevel mixed linear model* and HLM are exchangeable when the growth is supposed to be linear. *Mixed* here refers to the random and fixed effects contained in the models. Random effect is applied when the units such as children, classrooms, or schools are used to estimate the variance. For example, one asks whether some schools or school districts have bigger regression slopes or account for the variance in the outcome measure. Fixed effect generally refers to the grouping or treatment effects. In an observational study, it usually refers to the covariates at each level. Raundenbush (1988) indicated that in HLM where random parameters are included, estimation of fixed effects becomes more precise. Hence, the application of HLM can improve inference about fixed effects.
To enable an appropriate specification of the error structures, HLM considers the error structures as random coefficients, and thus improves the estimation of the variance (Raudenbush, 1988). HLM is a great improvement over the single-level, traditional approaches such as Ordinary Least Square (OLS) regression and Weighted Least Square (WLS) regression. Raudenbush & Bryk (1987) compared the two approaches with HLM model and showed that HLM provides a more efficient estimation and smaller standard errors. Goldstein (1987) also demonstrated how standard errors obtained with OLS can be misleading. Aitkin, Anderson, and Hinde (1981) confirmed it by comparing their multi-level analysis with the earlier analysis which had used a fixed-effect analysis. The reanalysis found similar magnitudes of effects, but larger standard errors when the random effect of classrooms was included. As a result, certain key hypotheses rejected in the original report were retained in the reanalysis. This application illustrated again the advantage of the HLM approach.

As to the application of HLM models, Raudenbush (1988) reviewed various studies where the applications of HLM models have enriched the study of variation between and within countries (Mason et al., 1984), schools (Wisenbaker & Schmidt, 1979; Raudenbush & Bryk, 1986; Lee, 1986; De Leeuw and Kreft, 1986; Goldstein, 1986; Aitkin and Longford, 1986), classrooms (Aitkin, Anderson, & Hinde, 1981), studies (Raudenbush & Bryk, 1985), and individuals (Bryk & Raudenbush, 1987; Ware, 1985). There also have been studies where HLM was applied to investigate the achievement growth (Muller, Stage, & Kinzie, 2001), and the ethnic differences in
achievement (Lee & Bryk, 1989; Raundenbush & Bryk, 1987). Without investigating the random effects of ethnicity across individuals and schools/districts, most of the studies only focused on the static difference between ethnic groups instead of seeing the achievement gap as ethnic differences in achievement growth. Meanwhile, according to Lee & Bryk (1989), very few studies investigated what factors are associated with the achievement gap.

**Contributions of the Study**

Basically, the dissertation expands the discussion from the following aspects:

(i) Debate over approaches

Very few studies compared the results of different approaches of investigating student achievement and school effectiveness. Only Raudenbush (1994) was found to compare the results of accountability systems based on student mean proficiency and those based on value added modeling.

The analyses conducted in this study aim to answer these questions: Do approaches matter? Does growth modeling give substantially different results from descriptive analyses when both are applied to investigate achievement gap? Do the association between school/district characteristics and achievement gap differ from the two approaches? If the traditional descriptive approach gives essentially the same results as the growth modeling, it is natural to argue on behalf of the simpler method. On the other hand, if the two show very different pictures of achievement gap and school effectiveness, researchers need to further address these differences. And in such a case, it
might be necessary for educators and researchers to report both approaches when the issue of achievement gap is involved.

(ii) Methodology advancement

The study seeks to provide a comprehensive understanding of the phenomenon of racial inequality in educational achievement from a growth perspective. The methodology applied here - HLM modeling, according to Singer & Willett (2003, Preamble vii), have yet to be “widely and wisely” used. Therefore, this study aims to contribute to the current discussion regarding the advances of the methodology in the field of education equity.

By tracking the same cohorts, the inter-group comparison becomes more convincing since the differences among cohorts can be excluded from the models applied in this study. The comparisons are then made between ethnic groups focusing on the achievement growth rates and intercepts instead of the achievement scores at one-time or multiple-time points. Conceptually, the specified growth model can be viewed as a within-person regression model representing individual change over time.

(iii) State-level instead of nation-level analysis

Although many studies and public discussion rely on data reported at national level, Raudenbush, Fotiu, Cheong, & Ziazi (1996) showed that there was more heterogeneity within states than among states. Patterns of achievement by race or trends in the achievement gap might present a different picture if they were viewed at lower levels of aggregation (i.e. state instead of national level) (Braun, et al, 2006). The dissertation focuses on one state’s data to examine the achievement growth differences.
(iv) Model-based methodology instead of simple descriptive statistics

Instead of using the descriptive methods such as percent achieving a performance standard or mean scale scores, the methodology used here is model-based. One limitation of the previous studies stems from the use of the fixed-effect models to estimate the achievement gaps. The models applied here provide a full view of the relationships of the variables at different levels so that not only can the fixed effects be examined; the random effects of individuals, schools and districts can also be investigated. Furthermore, investigating the relationship between initial status and rates of change helps broaden the research regarding school effects.

(v) Organizational effects of schools and school districts

The organizational effect draws our attention to the social system rather than individuals. The study takes into account the variation across individuals, schools, and districts and how they interact as a dynamic social system to affect racial inequality. The term organizational effect again refers to the relationship between achievement growth differences and the various demographic characteristics of schools and districts. It will only be used to explain whether the random difference among the organizations such as schools or districts may account for the variation of achievement growth. The word effect here contains no causal connotation and thus does not imply a causal relationship.

By expanding the growth model to a three-level HLM model, we can examine how the various school/district characteristics relate to differences among schools/districts in their mean rates of change, their initial statuses, and the interaction between the two. The study provides evidence with regard whether the school or district
characteristics are associated with the student achievement growth and the ethnic differences in achievement growth.

(vi) Focus on suburban schools

The study focuses on suburban schools and districts because (1) very few studies have investigated the achievement gap and the school segregation in suburban areas; (2) segregation is becoming more serious in suburban areas since more minority students live in such areas (Reardon & Yun, 2001).

(vii) Linking achievement gap with school segregation

According to Ipka (2003), very few papers were found to analyze the school segregation level when investigating the achievement gap. The degree of segregation is included as one of the district characteristics to carry out a thorough analysis regarding the relationship between the degree of segregation and the achievement growth differences.
Chapter 3: Research Design

“As interest in hierarchical models has grown, the pace of methodological innovation has accelerated, with many creative applications in social science and medicine” (Raudenbush & Bryk, 2002, p10).

Data and Measurement

The dissertation, in essence asks (1) How the outcome (student achievement) changes over time; and (2) Whether we can account for differences in these changes by identifying some relevant individual, school, and district characteristics. According to Singer & Willett (2003, p8), three important features of a study of change are:

- “Three or more waves of data”: The data used here contain three years of the assessment results (year 2002-03, 2003-04, and 2004-05). With only three waves of data, individual growth is usually assumed to be linear over time.
- “An outcome whose values change systematically over time”: The students’ mathematics achievement scores from the state-wide standardized tests change every year.
- “A sensible metric for clocking time”: Time is the fundamental predictor in the study of achievement growth. Naturally, grade and year is chosen as the time metric in this study (see Table 3.1).

The study focuses on one cohort of students listed above by tracking the growth in achievement scores, as well as by comparing the achievement growth among ethnic groups.
Data

Data Description

The data used in this dissertation are drawn from students from 6th grade to 8th grade in the years of 2002-03, 2003-04, and 2004-05 in 99 schools in 14 suburban public school districts in a single state. Of more than 15,000 students in the database, approximately 72 percent are White students, almost 19 percent are Hispanics students, around 4 percent are Blacks students (non-Hispanic), around 4 percent are Asian students, and about 1 percent of the students are of other racial and ethnic backgrounds. There are slightly fewer female students than male students, although their distributions in each ethnic group are quite similar. Table 3.2 shows the demographic information of the students whose test scores are to be investigated and compared in this study. For the two variables “gender” and “ethnicity”, there is no missing data since missing value in one year can be replaced by the data from the other two years.

The state’s target is 100 percent participation in the state assessment and its alternate assessment, both of which are linked to the state standards. The state Department of Education reports that less than 1% of the total student population is assessed through the alternate assessment at each grade and within each content area. However, some students are still excluded from these state assessments, including those
with disabilities who cannot receive accommodations to take the state test and the
English Language Learners who have not been continuously enrolled for one year. It was
reported that participation by students with disabilities in the 2003 math assessments
grew to 98.6 percent. In all, as the student samples in this study are compared to the
overall student population in the whole state, the white students are slightly over-
represented while the Hispanic students are slightly under-represented.

Table 3.2
Demographic Background of the Students in the Two Cohorts

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>619(4.2%)</td>
<td>338(2.3%)</td>
<td>2819 (1.9%)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>3004 (20.1%)</td>
<td>1510(10.1%)</td>
<td>1494(10.0%)</td>
</tr>
<tr>
<td>White</td>
<td>11286(75.7%)</td>
<td>5765(38.7%)</td>
<td>5521(37.0%)</td>
</tr>
<tr>
<td>Total</td>
<td>14909(100.0%)</td>
<td>7613(51.1%)</td>
<td>7296(48.9%)</td>
</tr>
</tbody>
</table>

The suburban schools are selected according to the information on the location
type taken from the National Center for Education Statistics (NCES), Common Core of
Data (CCD). The location type is a descriptive term used by the U.S. Department of
Education's National Center for Education Statistics (NCES) to indicate a district's urban,
suburban, or rural status, based on its location relative to populous areas. NCES has
developed eight locale type designations, which have been broken down as follows:

- Urban: Large Central City, Mid-Size Central City

- **Suburban** (Focus of the study): Urban Fringe of Large City, Urban Fringe of
  Mid-Size City, Large Town, Small Town

- Rural: Rural: Inside MSA (metropolitan statistical area), Rural - Outside MSA
In order to calculate a meaningful segregation index, schools and districts need to meet a series of criteria. Zoloth (1976) found that difficulties in interpretation may arise in two instances: “if the district contains only a few schools” and “if either the racial composition or actual degree of desegregation differs substantially between sets of schools offering different grade spans (e.g. between elementary and secondary schools)”. Since the study only focuses on the schools with a grade span from grade 6 to grade 8, the second criterion is not of concern. As to the first criterion, we exclude very small districts from the dataset to mitigate this problem. Therefore, the data only include those school districts which meet the following criteria: (1) It contained more than 3 schools; (2) at least 5 percent of the student population was minority; (3) at least 5 percent of the student population was non-minority. Note that two very large school districts which contain more than 50 schools from grade 5 to grade 7 were also excluded from the data. Table 3.3 presents the percentage of each racial group in the districts and how many schools a district has.

There are 239 students with no record of which districts they belong to. The racial distribution of the missing records is American Indian/Alaskan Native (3.3%), Asian/Pacific Islander (2.5%), Black (not Hispanic) (5.4%), Hispanic (20.5%), and White (68.2%). This distribution is close to the distribution of the whole sample. Hence, the missing cases will be neglected in the district-level analysis since they do not skew the whole distribution.

As to missing data, according to Singer & Willett (2003, p12), individual growth modeling does not require balanced data; in other words, each student’s growth record
can contain “a unique number of waves.” This approach allows considerable flexibility in estimation when the number and timing of the observations can vary over individuals (Laird & Ware, 1982; Strenio, Weisberg & Bryk, 1983).

Table 3.3
The Distribution of Racial Groups in the Districts

<table>
<thead>
<tr>
<th>No.</th>
<th>Sch .No</th>
<th>Native American</th>
<th>Asian</th>
<th>Black</th>
<th>Hispanic</th>
<th>White</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>24(9%)</td>
<td>130(5.2%)</td>
<td>64(2.4%)</td>
<td>630(25.9%)</td>
<td>1648(65.5%)</td>
<td>2496</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>6(7%)</td>
<td>10(1.6%)</td>
<td>14(1.9%)</td>
<td>228(41.8%)</td>
<td>332(54.0%)</td>
<td>590</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>5(6%)</td>
<td>67(10.1%)</td>
<td>13(2.2%)</td>
<td>365(51.3%)</td>
<td>255(35.8%)</td>
<td>675</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>4(2.4%)</td>
<td>5(1.5%)</td>
<td>11(5.0%)</td>
<td>60(24.2%)</td>
<td>153(66.9%)</td>
<td>233</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>17(.5%)</td>
<td>200(6.3%)</td>
<td>316(10.2%)</td>
<td>311(8.6%)</td>
<td>2481(74.4%)</td>
<td>3325</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>10(9%)</td>
<td>35(2.9%)</td>
<td>26(2.0%)</td>
<td>80(7.6%)</td>
<td>991(86.7%)</td>
<td>1142</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>11(9%)</td>
<td>48(3.1%)</td>
<td>17(1.1%)</td>
<td>387(25.2%)</td>
<td>1042(69.7%)</td>
<td>1505</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>17(.6%)</td>
<td>112(5.0%)</td>
<td>30(1.3%)</td>
<td>225(12.0%)</td>
<td>1751(81.1%)</td>
<td>2135</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>2(6%)</td>
<td>4(9%)</td>
<td>0(0%)</td>
<td>125(41.6%)</td>
<td>169(56.9%)</td>
<td>300</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>4(1.2%)</td>
<td>15(3.2%)</td>
<td>96(16.4%)</td>
<td>89(15.9%)</td>
<td>352(63.3%)</td>
<td>556</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>0(0%)</td>
<td>4(9%)</td>
<td>1(3%)</td>
<td>108(35.3%)</td>
<td>222(63.3%)</td>
<td>335</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>19(5.6%)</td>
<td>3(1.6%)</td>
<td>4(2%)</td>
<td>39(11.7%)</td>
<td>286(80.4%)</td>
<td>351</td>
</tr>
<tr>
<td>13</td>
<td>10</td>
<td>14(1.1%)</td>
<td>13(9%)</td>
<td>13(9%)</td>
<td>201(15.1%)</td>
<td>1158(82.0%)</td>
<td>1409</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>4(8%)</td>
<td>1(6%)</td>
<td>1(8%)</td>
<td>107(23.1%)</td>
<td>283(74.7%)</td>
<td>396</td>
</tr>
<tr>
<td>Missing</td>
<td>8(3.3%)</td>
<td>6(2.5%)</td>
<td>13(5.4%)</td>
<td>49(20.5%)</td>
<td>163(68.2%)</td>
<td>239</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>145(9%)</td>
<td>653(4.2%)</td>
<td>619(3.9%)</td>
<td>3004(19.1%)</td>
<td>11286(71.9%)</td>
<td>15707</td>
<td></td>
</tr>
</tbody>
</table>

For the purpose of this study, the records of the students with at least two scale scores are included to construct a projection. Therefore, for the students who are only linked to two scale scores in the dataset, the growth modeling can still be applied with
one-year score missing. Non-random missing data, however, is problematic for drawing inferences. Cases with missing values which are systematically different from cases without missing values can result in biased estimates. In order to decide whether or not data are randomly missing, the associations between missing data and other variables will be described. Decisions will then be made on whether to exclude or impute the missing value as well as on how to carry out the imputation if the missing data are included.

Given that HLM does not allow missing values on upper-level variables (i.e. at level two (individual) or level three (school or district)), the missing values will be replaced or imputed at these levels. The rule by Arnold (1992) will be applied that variables with more than 20% missing data are not included in the analyses. The numbers of students, schools, and districts are subject to change for the final analysis after further data cleaning and taking into account the factors such as missing data, student mobility, etc. For different analyses, the number of units may vary.

Another thing to be checked is the student IDs. The student IDs must uniquely identify students. The duplicate IDs were checked and eliminated so that each student has at most one record per year per subject. Some cases with duplicate records had a record with a valid scale score and another record with missing data. The cases with the missing data were removed, leaving one record with a valid scale score.

- Those students with duplicate valid records were removed.
- Those students with duplicate records with two missing scale scores were included in the analysis. We “randomly” deleted the second of the duplicate
records leaving a single record for the student in the given subject by year combination with a missing scale score.

Therefore, students who were NOT included in the analyses include:

(i) Students who were not in suburban school districts

(ii) Students without two mathematics scale scores with which to construct a projection.

(iii) Students with a bizarre sequence of grades (see Appendix A).

Measurement

The variables are constructed from two major sources. The first is a state-wide student assessment program, which is designed to provide a picture of how students in the state progress toward meeting academic standards, and how schools perform to ensure learning success of students. The second source of the measures is the Common Core of Data (CCD), from which we will identify and combine some other factors at school or district level such as percentage of the students who received free or reduced-price lunch, teacher-pupil ratio, etc.

The student assessment program is a large-scale standardized paper–pencil achievement test administered every year. The purpose of the state assessment is to provide an annual measure of student performance relative to the state content standards. All students in grades 3 through 10 sit for Reading/Writing and Mathematics and students in grades 5, 8, and 10 sit for Science. There is a single form for each grade. All forms are timed assessments administered under standardized conditions to support the reliability
and validity of the test results. The state assessments were developed by CTB/McGraw-Hill in collaboration with the Department of Education and were scored and scaled by CTB/McGraw-Hill.

Students’ total scale scores are based on their performance on all the scored items on the test. Students also receive a score for each sub-content area that is based only on the items that contribute to the given content standard (or sub-content area). Note that every item on the test corresponds to some content standard but not all items contribute to a sub-content area. Students were scored at the total test, content standard, and sub-content area levels using item response theory pattern scoring procedures. This procedure produces maximum-likelihood trait estimates (scale scores) based on students’ item response patterns, as described by Lord (1974; 1980, pp. 179-181). Item pattern scoring takes more information into account and is more accurate than number-correct scoring in which all students with the same number correct receive the same score, regardless of how that score is obtained. Stated differently, item pattern scoring takes into account the information on how the scores are obtained instead of simply summing up the number of correct items.

CTB uses item response theory (IRT) to place multiple-choice and constructed response items on the same scale. Because the characteristics of selected response (multiple-choice) and open-ended response (constructed-response) items are different, two item response theory models are used in the analysis of test forms containing both item types. The three-parameter logistic (3PL) model is used for the analysis of selected-

---

2 Part of the section of measurement is drawn from the technical report by CTB/McGraw-Hill.
response items. For analysis of constructed-response items, the two parameter partial credit model is used.

Reliability is an index of the consistency of test results. A reliable test is one that produces scores that are expected to be relatively stable if the test is administered repeatedly under similar conditions. The estimated reliability index (Cronbach’s alpha), total score reliability coefficients, for the total test and for each content standard at each grade are all greater than .85. These coefficients tend to be somewhat lower than the coefficients for the total test scores.

The range of possible scores varies by grade and content area. Continuity of the test results between years within the same grade and between grades (vertical scale) is maintained using both an "anchoring" of items within tests and shared items between grades. The lowest scores for grades 6 to 8 are 240, 280, and 310 while the highest scores for the three grades are 830, 860, and 890. The scale scores are all aligned in a continuous scale.

Student results provide valuable information used to determine longitudinal growth. In addition to measuring the standards as delineated in assessment frameworks, the assessment measures the progress of students over time. With the development of vertical scales, the progress of each student and group of students can be examined each year.

In order to measure change, the outcome measures should be equated over time so that the scores can be equivalent across years. Equating is necessary to account for slight differences in test difficulty and maintain scale comparability across administrations.
Each form of a new test includes a subset of items used in the previous administrations of the assessments. These repeated items are used to equate the forms across years. The within-grade equating is used to account for year-to-year differences in test difficulty and to maintain comparability across years. In order to equate current tests to base year scale, a set of multiple-choice anchor items was selected for each grade in Mathematics. These items demonstrated good classical and IRT statistics and represented the test blueprint.

By having common items between adjacent grades, the vertical scales have been established so that the unique metrics of the mathematics vertical scales were maintained across grades. The Stocking and Lord (1983) procedure was used for each grade. For example, Grade 7 was linked to grade 8, and grade 6 was subsequently linked to grade 7. For the same grade, each year’s test contained items from the previous administrations. These repeated items were used as anchors in a Stocking and Lord (1983) equating procedure. Thus, the scale for each test form was linked to the previously established vertical scale.

Horizontal equating within each grade was used to place the new forms on the vertical scales that had been established previously for the subject. The vertical scale for Mathematics, spanning grades 5 through 10, was established in 2002. The Stocking and Lord (1983) procedure was used to place each grade on the vertical scale that had been developed for each content area. Each test contained items from the previous administrations for the same grade. These repeated items were used as anchors in a Stocking and Lord (1983) equating procedure, which was used to place each test form on
the previously established scale. By equating the tests at a certain year within each grade, the unique metrics of the Mathematics vertical scales were maintained.

Once a vertical scale has been chosen, the scaling procedure itself may have an impact on the growth trajectory patterns. As discussed in the literature review, some studies found that applying different scaling methods can lead to different conclusions about the variance in student achievement scores. Ambiguity does exist when the question arises—had the test publisher used another scaling method, how would the results change? The debate over the choice of scaling method at the end of 1990’s did not conclude with a consensus on the best approach. The assessment results in this state have been built on the IRT scaling method, which may provide a limited scope of achievement growth and achievement growth gap. However, the investigations, based on particular designs and assumptions, can still lead us to a better understanding of racial inequality in educational achievement in today’s public schools and districts. Future studies are needed to further investigate the impact of different scaling methods on measurement of growth.

Another assumption worthy of mention is that the measures of student achievement are on interval scales. The interval nature of scale refers to the assumption that a difference of 1 point, wherever it occurs on the scale, means the “same thing” in terms of magnitude of difference with regard to the construct under consideration. When achievement growth is projected, this assumption indicates that the locations of scores have no impact on the growth trajectories. This assumption is to be examined by including the relationship between growth intercept and growth rate into the analysis.
Variables

**Dependent Variable**

For the research question—have the achievement scores of minority students (Black students and Hispanic students) grown faster on average than the scores of White students, we apply the two-level conditional HLM model. Interest centers on the achievement growth parameters, *the growth intercept and growth rate*.

For the research question—do the differences in achievement growth between ethnic groups vary across schools or districts and, if yes, which school or district factors are associated with such variation, we apply the three-level conditional HLM model. The parameter of interest is *difference in achievement growth among ethnic groups*. The parameter is related with the random effects of schools/districts, meaning that schools/districts may contribute to part of the variation of the achievement growth gap. Raundenbush & Bryk (2002, page 127) added the random effects of schools into the HLM model when they were trying to explain how the achievement gaps differ from school to school.

**Student-Level Predictors**

Among the student characteristics considered are gender and race/ethnicity. Race/ethnicity is a very important predictor here since the term “achievement growth difference” in the dissertation refers to the differences in the average achievement growth between minority and white students.

The interaction between gender and ethnicity will be investigated at this level.
Student mobility will also be considered. In the three academic years, students have the possibility of moving once or twice. The school codes across three years are used to designate a student as mobile or not mobile. For example, if a student’s school code was changed from year 1 to year 2, it means that this student moved between the two years. However, due to the limited grade span in each school, children may have changed schools because they reached the end of the grades offered at their previous schools. Thus, it will be very difficult to distinguish between those who "moved" and those who switched schools for the reason of “bureaucratic transition”. Instead of simply deleting the schools which do not have the targeted grade spans (from grade 6 to 8), different indices are given to identify different patterns of student mobility (see table 3.4).

Table 3.4  
*Student Mobility Index for Different Patterns of Moving*

<table>
<thead>
<tr>
<th>Mobility Index</th>
<th>Mobility Pattern</th>
<th>Bureaucratic move (year 1 &amp; 2)</th>
<th>Move (year 1 &amp; 2)</th>
<th>Bureaucratic move (year 2 &amp; 3)</th>
<th>Move (year 2 &amp; 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0010</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0100</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0001</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1001</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0110</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0101</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Note that mobility index 0 indicates that students did not move during the three years; index 1 indicates that students moved once because they reached the end of the grades offered in their schools (bureaucratic transition); index 2 indicates that students
moved once; index 3 indicates that students moved twice and that one of the two moves was due to bureaucratic transition; index 4 indicates that students moved twice.

**District- or School-Level Predictors**

*The measures of segregation*

In order to investigate the relationship between the degree of segregation and the student achievement growth differences, a number of segregation indices have been reviewed and the segregation index $H$ was selected. The Information Theory Index ($H$), is also called Entropy Index in Reardon & Yun (2001) since Theil and Finizza (1971) offer the interpretation of $E_i$ as the “racial entropy” of the student body of the $i$th school. While Entropy by itself is not a measure of segregation, it is a useful index to summarize the overall race/ethnic mix in a school.

The formula for $H$ (Theil & Finezza, 1971) starts from the calculation of Entropy index ($E_i$). For the $i$th school

$$E_i = \sum_{m=1}^{M} p_{im} \ln \left( \frac{1}{p_{im}} \right)$$

where $p_{i}$ is the proportion minority of the $i$th school, and $M$ is the number of the minority groups. When $p_{im} = 0$, it means that there is no ethnic group $m$ (i.e. Black) in a particular school. Thus, this ethnic group will not be included in the calculation.

The entropy can be seen as a measure of “diversity” where the value of zero indicates that all individuals are members of a single group (“no diversity”), and the value of one indicates that individuals are evenly distributed among the $M$ groups.
\( p_{im} = 1/M \) for all \( i \). The maximum score is given by the natural log of the proportion minority in the calculation, but this maximum is only achieved when all groups have equal representation, which is seldom the case.

For the district as a whole, the entropy index \((E)\) is:

\[
E = \sum_{m=1}^{M} p_m \ln\left(\frac{1}{p_m}\right)
\]

Then we define \( H \) as:

\[
H = \frac{E - \sum_i \frac{T_i}{T} E_i}{E}
\]

where \( T_i \) and \( p_i \) are, respectively, the total enrollment and proportion minority of the \( i \)th school, and where \( T \) and \( p \), respectively, are the total enrollment and proportion minority of the district. \( p_m \) is the proportion in group \( m \) (e.g., proportion Black).

Another formula for \( H \) (Theil, 1972; Theil & Finezza, 1971) is

\[
H = \sum_{m=1}^{M} \sum_{i=1}^{L} \frac{T_i}{TE} p_{im} \ln \frac{p_{im}}{p_m}
\]

where \( p_{im} \) is proportion in group \( m \) in school \( i \) (e.g., proportion black in school \( i \)).

The value of \( H \) ranges from 0 to 1, where 0 indicates that all schools have identical racial composition and hence are equally “diverse”, and 1 indicates that each school is monoracial. Theil and Finizza (1971) have directly derived the index as a measure of school segregation. They stated that essentially \( H \) is a measure of “how much less diverse individual schools are, on average, than their district as a whole.” Therefore, it is only an indicator of the relative degree of segregation, referring to the weighted
average of each school’s deviation from the ethnic diversity of the overall district. This measure does not only depend on the race/ethnic composition of the population but also contains the information on how evenly population groups are distributed among schools or neighborhood.

Since no single index may be sufficient because of the complexity of the notion of segregation, the measures of racial composition of schools or districts are included in the analyses to obtain a different perspective on racial distribution. The measures generally consist of the two parts: the percentage of minority (non-white) students and the percentage of students in a specific minority group (i.e. Hispanic students only). Since the multi-group H index has taken into account the proportion of each minority group in schools and districts, the calculation of the percentage of minority students is only based on the cohort. In sum, the percentage of students in each minority group is not included (1) to avoid the redundant information since the segregation index is derived from it; (2) the $H$ index contains more information by producing a measure of school segregation across each district.

Therefore, the measures of district segregation include:

- The percentage of minority for the cohort
- The Information Theory Index ($H$)

School-level segregation typically refers to the racial segregation of students between classrooms within grade level, which is not the focus of the study. However, the measures of racial composition of schools are to be applied to indicate the degree of
racial diversity at the school level. The specific percentages of both Black students and Hispanic students are included.

*The measure of cohort mobility*

In the analysis of a longitudinal dataset, student mobility needs to be taken into account. As we discussed in the section on student-level variables (p.50), students can be identified as non-mover, bureaucratic mover, and mover. The school mobility rate takes into account both the number of students who left a school and the number who newly enrolled. These numbers are summed and then divided by the total number of students attending the school. Since we only focus one cohort of students by tracking their achievement growth, the mobility of the cohort will then be considered instead of school mobility. Therefore, the numbers in the formulas below are only based on the grades in the two cohorts.

The formula for the cohort mobility rate is:

$$ Mobility\_Rate = \frac{N_{in} + N_{out}}{Enrollment_{year\_2}} $$

$N_{in}$ is the number of students in the cohort who newly enrolled, then an in-mobility rate refers to the percentage of students who newly enrolled.

$N_{out}$ is the number of students in the cohort who left the school, then an out-mobility rate refers to the percentage of students who left the school.
Enrollment_{year 2} is the total number of the students in the cohort who enrolled in the school in the second year of the three, which is year 2003-04 (year 2\(^3\)).

**Other Variables**

Several other variables are also included in order to investigate what school or district characteristics are related to individual achievement growth and ethnic differences in achievement growth. They are:

- School/district size (school/district information from the CCD)
- Percentage of the students who receive free or reduced-price lunch (school/district information from the CCD)
- Pupil/teacher ratio (school/district information from the CCD)
- Percentage of special Ed/IEP students (school/district information from the CCD)
- Percentage of ELL (English Language Learner) students (school/district information from the CCD)

**Analytic Strategy**

This study of student growth involves a doubly nested structure of repeated observations (math achievement scores through three academic years) within individuals (students), who are in turn nested within organizational settings (schools or districts). The dissertation will focus on the continuous outcome - student achievement - and ask how

---

3 Year 1 refers to the first academic year in this dataset, year 2003-04; year 2 refers to the second academic year, year 2004-05; and year 3 refers to the third academic year
this variable changes over time; that is, how student achievement varies as a function of time and other predictors.

**Descriptive Analyses**

The analyses will proceed in two phases. The first is descriptive and the second is model-based. In the first phase, we will investigate whether the achievement gaps in different grades have decreased through the three years from 2002 to 2005.

**Achievement Gaps**

Both the mean scale scores and the variance in scores of different ethnic groups will be examined to investigate the trend in the achievement gaps. The achievement gap in this dissertation does not only refer to the difference in the mean scale scores, but also the spread in the distribution of scores. Hence, in addition to mean scale score, variance of scores is treated as another indicator of education equality. The variance of scores of different ethnic groups will be examined to give a more comprehensive picture of the achievement gap. Moreover, the effect size will be applied to further investigate the achievements gaps between different ethnic groups.

The possible comparisons include:

- Whether the gap in the mean scale scores between White and Black OR between White and Hispanic students observed in Grade 6 has increased or decreased through the three academic years;
- Whether the variance of the scale scores of students of different ethnicities in Grade 6 have increased or decreased through the three academic years
In addition to the general trend in achievement gaps, for which the calculation is based on the overall achievement scores, we will also address the question—whether the average achievement scores and the achievement gaps of different schools and school districts differ from each other. The average score and the achievement gaps will be compared at the three levels: school, district, and overall.

*Mobility*

The following descriptive analysis of mobility will be also presented:

- Characteristics of mobile students (ethnicity, gender, and grade)
- Description of the cohort mobility index
- Correlation between the cohort mobility rates and the achievement gaps at the school level

*Segregation*

We will investigate the degrees of the segregation in schools or school districts. The racial composition of schools or districts will be displayed, and the segregation index will be calculated to offer a better sense of the relative degree of segregation in public school system. The following descriptive analysis of segregation will be also presented:

- Correlation between the degree of segregation and other district characteristics
- Correlation between the degree of segregation and the achievement gaps at the district level

*Further Analysis*

- Correlation between the achievement gaps and other school/district characteristics.
• Description of average achievement scores and average achievement gaps at the district level
• Description of average achievement scores and average achievement gaps at the school level

Hierarchical Linear Modeling

The second phase relies on Hierarchical Linear Modeling (Raudenbush & Bryk, 2002) to investigate the growth of the individual learner within the organizational context of schools or school districts. To study the growth of students who are nested within schools, individual growth trajectories comprise the level-1 model; the variation in growth parameters among students within a school is captured in the level-2 model; and the variation among schools or districts is represented in the level-3 model.

Two-Level Model

Specifically, the two-level HLM model used is called Repeated-observations Model to address questions about change. According to Raudenbush & Bryk (2002, preamble vi.), there have been a number of other studies about change—“individual growth modeling (Rogosa, Brandt, & Zimowski, 1982; Willett, 1988), multilevel modeling (Goldstein, 1995), hierarchical linear modeling (Raudenbush & Bryk, 2002), random coefficient regression (Hedeker, Gibbons, & Flay, 1994), and mixed modeling (Pinheiro & Bates, 2000).”

We will apply a two-level HLM model first in order to investigate the achievement growth differences without considering the organizational effects of schools or districts.
Two-Level Unconditional Model

The research question to be answered by this model is—does the achievement growth over the three years vary across individuals?

We will first present an unconditional model to give an overview of the model structure. Then we will add covariates to investigate the association between student ethnicity with the achievement growth differences when holding other student-level predictors constant.

The unconditional model is also a “null” model with no independent variables, at either the individual or the school/district level. It is used to partition the variance in the outcome measure (mathematics scores) into within- and between-individual components. We begin at level 1 with an individual growth model of academic achievement at time $t$ of student $i$:

$$ Y_{it} = \pi_{0i} + \pi_{1i} (ACADEMIC.YEAR)_{it} + e_{iti}, \quad e_{iti} \sim N(0, \sigma^2) $$

where

- $Y_{it}$ is the math achievement score at time $t$ for student $i$;
- $(ACADEMIC.YEAR)_{it}$ is 0 for the 2002-03 academic year, 1 for 2003-04, 2 for 2004-05;
- $\pi_{0i}$ is the initial status of student $i$, that is, the expected outcome for that student in the academic year of 2002-03 (when ACADEMIC.YEAR=0);
- $\pi_{1i}$ is the growth rate for student $j$ during the three academic years; and
\( e_{ni} \) is the residual variance at level 1 after controlling for the academic year. It is assumed independently distributed with distribution \( N(0, \sigma^2) \) with mean 0 and variance \( \sigma^2 \).

The results of a preliminary analysis will be presented to suggest considerable random variation in \( \pi_0 \) and \( \pi_1 \) at level 2, the person-level model.

Specifically, at level 2, a between-person model,

\[
\begin{align*}
\pi_{0i} &= \beta_{00} + r_{0i}, \\
\pi_{1i} &= \beta_{10} + r_{1i},
\end{align*}
\]

where

\( \beta_{00} \) is the mean initial status (intercept)

\( \beta_{10} \) is the mean growth rate over three years

\( r_{0i} \) and \( r_{1i} \) are assumed to be multivariate normally distributed, both with expected values of 0. We label these variances as

\[
\begin{pmatrix}
\tau_{00} \\
\tau_{11} \\
\tau_{01}
\end{pmatrix}
\]

and the covariance between them as

\[
Cov(r_{0i}, r_{1i}) = \tau_{01}
\]

Putting them into a variance-covariance matrix

\[
Var(\pi) = \begin{pmatrix}
\tau_{00} & \tau_{01} \\
\tau_{10} & \tau_{11}
\end{pmatrix}
\]

Two-Level Conditional Model

The research questions to be answered by this model are - (1) Have the achievement scores of minority students (Black students and Hispanic students) grown
faster on average than the scores of White students? (2) To what extent can these differences be accounted for by student characteristics?

We now consider a model that allows estimation of the effects of student ethnic background and other variables of student level. The level-1 model remains the same as the one in the unconditional model. The level-2 model represents the variability in each of the growth parameters, \( \pi_{pi} \) (p=0, 1), among students within schools or districts. The fixed effects of student ethnic background will be represented here. Ethnic differences in student achievement can vary randomly across schools (i.e. in different schools the student achievement gaps might be different). Thus, the coefficient of a dummy variable for ethnicity can be used to examine the random effects of schools or districts in the three-level model presented later.

Specifically, for the achievement growth differences between Black and White students, the level-2 model is formulated as follows:

\[
\pi_{0i} = \beta_{00} + \beta_{01}(ETHNICITY_{B-W})_i + \sum_{k}^{K} \beta_{0k}X_{ik} + r_{0i}
\]

\[
\pi_{1i} = \beta_{10} + \beta_{11}(ETHNICITY_{B-W})_i + \sum_{k}^{K} \beta_{1k}X_{ik} + r_{1i}
\]

Or for the achievement growth gap between Hispanic and White students

\[
\pi_{0i} = \beta_{00} + \beta_{01}(ETHNICITY_{H-W})_i + \sum_{k}^{K} \beta_{0k}X_{ik} + r_{0i}
\]

\[
\pi_{1i} = \beta_{10} + \beta_{11}(ETHNICITY_{H-W})_i + \sum_{k}^{K} \beta_{1k}X_{ik} + r_{1i}
\]

where \( \beta_{00} \) represents the mean initial status
\( \beta_{10} \) is the mean growth rate over the three academic years

\( \beta_{01} \) is the differences in intercept (initial differences) between white and minority students in the first academic year (2002-03), holding other covariates constant

\( \beta_{11} \) is the differences in the growth rates between white and minority students, holding other covariates constant

\( X_{ik} \) represents other predictors when \( K \) is the number of the predictors

\[
( X_{i2} \ldots X_{ik} )
\]

\( \beta_{0k} \) represents the direction and strength of association between the predictors \( X_{ik} \) and \( \pi_{0i} \) the intercept.

\( \beta_{1k} \) represents the direction and strength of association between the predictors \( X_{ik} \) and \( \pi_{1i} \) the growth rate.

\( r_{0i} \) and \( r_{1i} \) are assumed to be multivariate normally distributed, both with expected values of 0. The variance decomposition from a conditional model with more than one predictor can be checked by chi-square to see whether the reduction in variance is significant given the degree of freedom.

The person-level predictors here include student mobility and gender. The differences in achievement growth between male and female students as well as between students who moved and who did not move will be investigated. Meanwhile the two predictors serve as covariates so that we can detect the ethnic differences when holding student gender and mobility constant.

It is likely that minority groups have lower starting points and thus tend to grow faster (i.e. regression toward the mean) than their non-minority counterparts. Regression
toward the mean may need to be controlled for. Our solution is to investigate the relationship between the initial status and the rates of change.

When the initial status is included into the model, it will be grand mean centered. In general, the choice of centering at upper level (the second level in this model) is not as critical as the one in the first level (Raudenbush & Bryk 2002, p35). Variables at upper levels in HLM models can be un-centered or grand mean centered. Variables at lower levels can be un-centered, group mean centered, or grand mean centered.

**Three-Level Model**

The three-level model takes into account the possibility of random variation across schools or districts. In the dissertation, we generally refer to it as organizational effect\(^4\). This effect rests on an assumption that students will attend a specific school or district for a sufficiently long period of time so that the school or district can make a difference in their achievement. The assumption makes it necessary for us to examine mobility at the student and school levels. The students who changed schools during the three academic years have more than one value for their school or district variables. Instead of simply deleting the records of the mobile students, we average these variables (i.e., school size, segregation index, etc). Each student will then have a single value for each school-level variable.

The three-level HLM will be applied to partition overall achievement variance into within-student, between-student, within-school/district, and between-school/district

\(^4\) Note that the term *organizational effect* does not imply causal effect. It only refers to how much variation of achievement growth can be accounted for by the random variation across schools or districts.
variance components. The basic variance component model will later be connected with school/district characteristics for the third-level of HLM.

Schools and school districts will be put in separate but parallel models. Hence, the two parallel models are (1) test scores (first level)-student (second level)-school (third level); (2) test scores (first level)-student (second level)-district (third level).

**Three-Level Unconditional Model**

The research question to be answered by this model is—do the ethnic differences in achievement growth vary substantially across schools or districts?

The unconditional model is first presented to give an overview of the model structure. We begin at level 1 with an individual growth model of the academic achievement at time $t$ of student $i$ in school $j$:

$$Y_{ij} = \pi_{0ij} + \pi_{1ij} (ACADEMIC\ YEAR)_{ij} + e_{ij}, \quad e_{ij} \sim N(0, \sigma^2)$$

where

- $Y_{ij}$ is the math achievement score at time $t$ for student $i$ in school/district $j$ ($j$ here indicates schools or districts)
- $(ACADEMICYEAR)_{ij}$ is 0 for the 2002-03 academic year, 1 for 2003-04, 2 for 2004-05
- $\pi_{0ij}$ is the initial status of student $ij$, that is, the expected outcome for that student in the academic year of 2002-03 (when ACADEMICYEAR=0)
- $\pi_{1ij}$ is the growth rate for student $ij$ during the three academic years
- $e_{ij}$ is the residual variance at level 1 after controlling the for the academic year. It is assumed independently distributed in the distribution $N(0, \sigma^2)$ with mean 0.

The results of a preliminary analysis will be presented to suggest considerable variation in $\pi_0$ and $\pi_1$ at both level 2 and 3.
Specifically, at level 2, which is the person-level model,

\[ \pi_{0ij} = \beta_{00j} + r_{0ij}, \]
\[ \pi_{1ij} = \beta_{10j} + r_{1ij}, \]

and, at level 3, which is called school-level or district-level model (school or district are the third level in our separate models),

\[ \beta_{00j} = \gamma_{000} + u_{00j}, \]
\[ \beta_{10j} = \gamma_{100} + u_{10j}, \]

Where

\( \beta_{00j} \) represents the mean initial status within school/district \( j \)

\( \gamma_{000} \) is the average mean initial status across schools/districts

\( \beta_{10j} \) is the mean growth rate of three academic years within school/district \( j \)

\( \gamma_{100} \) is the average mean growth rate of three academic years across schools/districts

For level two, we assume that \( r_{0ij} \) and \( r_{1ij} \) are multivariate normally distributed, both with expected values of 0. We label the variances as

\[ Var(r_{0ij}) = \tau_{\pi11}, \]
\[ Var(r_{1ij}) = \tau_{\pi22} \]

and the covariance between them as

\[ Cov(r_{0ij}, r_{1ij}) = \tau_{\pi12} \]

Collecting these terms into a variance-covariance matrix

\[ Var_{\pi} = \begin{pmatrix} \tau_{\pi11} & \tau_{\pi12} \\ \tau_{\pi12} & \tau_{\pi22} \end{pmatrix} \]
For level three, we assume that $\mu_{0ij}$ and $\mu_{1ij}$ are multivariate normally distributed, both with expected values of 0. We label these variances as

$$\text{Var}(\mu_{00j}) = \tau_{\beta11},$$

$$\text{Var}(\mu_{10j}) = \tau_{\beta22}$$

and the covariance between them as

$$\text{Cov}(\mu_{00j}, \mu_{10j}) = \tau_{\beta12}.$$

Collecting these terms into a variance-covariance matrix

$$\text{Var}_\beta = \begin{pmatrix} \tau_{\beta11} & \tau_{\beta12} \\ \tau_{\beta12} & \tau_{\beta22} \end{pmatrix}.$$

**Three-Level Conditional Model**

The research question to be answered by this model is—Do the differences in achievement growth between ethnic groups vary across schools or districts, and if yes, which school or district factors are associated with such variation?

We now consider a model that allows estimation of the effects of student ethnic background, as well as of other characteristics of students and schools/districts. Moreover, among all the school- or district-level variables, the predictions of the segregation on the achievement growth differences can be investigated when other predictors are held constant.

The level-1 model remains the same as the one in the unconditional model. The level-2 model represents the variability in each of the growth parameters, $\pi_{psi} (p=0, 1)$, among students within schools or districts. Again, at the level-2 model, we can compare the achievement growth between different ethnic groups when the variable ethnicity is
included. More importantly, in this model we can investigate the random effect of schools and districts, where the achievement growth and ethnic differences in achievement growth can vary significantly.

To model the variances, the level-2 model is formulated as follows:

\[
\pi_{0ij} = \beta_{00j} + \beta_{01j} (ETHNICITY_{B-W})_{ij} + \sum_{k} \beta_{0kj} X_{ikj} + r_{0ij}
\]

\[
\pi_{1ij} = \beta_{10j} + \beta_{11j} (ETHNICITY_{B-W})_{ij} + \sum_{k} \beta_{1kj} X_{ikj} + r_{1ij}
\]

Or for the achievement growth gap between Hispanic and White students

\[
\pi_{0ij} = \beta_{00j} + \beta_{01j} (ETHNICITY_{H-W})_{ij} + \sum_{k} \beta_{0kj} X_{ikj} + r_{0ij}
\]

\[
\pi_{1ij} = \beta_{10j} + \beta_{11j} (ETHNICITY_{H-W})_{ij} + \sum_{k} \beta_{1kj} X_{ikj} + r_{1ij}
\]

Where

\(\beta_{00j}\) is the mean status in school j for a Black/Hispanic student

\(\beta_{01j}\) is the achievement gap between Black and White students or between Hispanic and White students on initial status, holding other covariates constant

\(\beta_{10j}\) is the 3-year growth rate for a Black/Hispanic student in school/district j

\(\beta_{11j}\) is the achievement gap on the 3-year growth rate in school/district j

\(X_{ikj}\) represents other predictors when k is the number of the predictors \((X_{i2j} X_{ikj})\);

\(\beta_{0kj}\) represents the direction and strength of association between the predictors \(X_{ikj}\) and \(\pi_{0ij}\) the intercept; and

\(\beta_{1kj}\) represents the direction and strength of association between the predictors \(X_{ikj}\) and \(\pi_{1ij}\) the growth rate.
The level-3 model represents the variability among schools/districts in the $\beta$ coefficients. Here is the following level-3 model:

$$
\begin{align*}
\beta_{00j} &= \gamma_{000} + \sum_{q} \gamma_{001} X_{qj} + \mu_{00j} \\
\beta_{01j} &= \gamma_{010} + \sum_{q} \gamma_{011} X_{qj} + \mu_{01j} \\
\beta_{10j} &= \gamma_{100} + \sum_{q} \gamma_{101} X_{qj} + \mu_{10j} \\
\beta_{11j} &= \gamma_{110} + \sum_{q} \gamma_{111} X_{qj} + \mu_{11j}
\end{align*}
$$

where

$\gamma_{000}$ is the average mean initial status for minority students across schools/districts.

$\gamma_{010}$ is the average intercept difference between white and minority students across schools/districts.

$\gamma_{100}$ is the average mean growth rate of three academic years across schools/districts.

$\gamma_{110}$ is the average growth rate difference between white and minority students across schools and districts.

$X_{qj}$ represents the other predictors at the school/district level when $q$ is the number of the predictors ($X_{1j} \ldots X_{qj}$).

$\gamma_{001}$ is the intercept for predictor q at school/district level.

$\gamma_{011}$ represents the direction and strength of association between the predictors $X_{qj}$ and $\pi_{1q}$ the growth rate at school/district level.

$\gamma_{101}$ is the mean growth rate for predictor q at school/district level.

$\gamma_{111}$ represents the direction and strength of association between the predictors $X_{qj}$ and $\pi_{1q}$ the growth rate at school/district level.
Since the three-level conditional model contains more than one predictor, the variance decomposition will then be checked by chi-square to see whether residual parameter variance still remains to be explained.

**Future Results and Discussion**

In sum, in addition to presenting the descriptive results regarding the achievement gaps, segregation, and mobility, the following HLM model-based results will be presented and discussed:

- The relations of person-level and school/district-level predictors to both status and change
- The estimated variances and related statistics
- The relationship between the initial statuses and rate of change

Other technical details will be investigated and discussed. They include:

- The centering of predictors
- The interaction effects between the predictors

Finally, the results from the descriptive analyses will be compared with the results from the Hierarchical Linear Models. The terms *achievement gap* and *ethnic difference in achievement growth* will then be examined and compared. The comparison between the two different approaches aims to answer the question—does methodology really matter while investigating the achievement gaps? Or to be more specific, does growth modeling bring us a new perspective on the issue of racial inequality in educational achievement?
The new term *ethnic differences in achievement growth* will then support the further discussion regarding potential policy change.
Chapter 4: Results

“No one method is necessarily considered superior; but each has strengths and shortcomings that researchers should be aware of in order to select the analytic method best suited for the particular research context” (Hancock & Lawrence, 2006, p171).

The analyses conducted for the dissertation, as described in the chapter on research design, can be divided into two parts. The first part involves the descriptive analyses examining the trends in achievement and achievement gaps, as well as the relevant independent variables at student, school and district levels. The second is the application of Hierarchical Linear Modeling (HLM) analyzing the ethnic differences in achievement growth and the organizational effects\(^5\) of schools and school districts. By focusing on the achievement gaps between Hispanic and White students as well as between Black and White students, all these analyses were only based on three ethnic groups—Black, Hispanic, and White in the state mathematics assessment system. The purpose of conducting different analyses is to illustrate various perspectives on ethnic differences in student achievement when achievement gaps and achievement growth gaps were both investigated. The comparisons between different approaches lead to the discussion about whether methodology matters in the investigation of achievement gaps. The analyses in both parts were based on the achievements of the student cohort of grade 6 in 2002-03, grade 7 in 2003-04, and grade 8 in 2004-05.

\(^5\) Note that the term *organizational effect* does not imply causal effect. It only refers to how much variation in achievement growth can be accounted for by the variation across schools or districts.
Section One: Descriptive Analyses

Mean scale scores and standard deviations are presented to describe the trends in achievement and achievement gaps. Effect size is also applied as the standardized difference between the means of different ethnic groups. Gain score is used as the principal indicator of achievement growth. Other than describing the trends, this study identifies what factors may account for the variation of achievement and achievement gaps. Student demographic background and school/district characteristics are examined and further linked to achievement and achievement gaps. Among all the independent variables, index of cohort mobility and index of segregation are the two very important ones in the longitudinal analysis regarding racial differences in educational achievement.

Achievement and Achievement Gaps

The descriptive analyses of achievement and achievement gaps can be conducted at the three levels—overall (student), school and district. At the overall level, the calculations of achievement and achievement gaps were based on all the scores of students while at the two other levels, the calculations were based on the aggregated information (average scores) of schools and districts. The approaches at different levels provide a comprehensive view of student achievement and achievement gaps in the suburban public schools.

Overall Level

Scale score

There were 14905 student records for the cohort in this study. The numbers of Black, Hispanic, and White students are 617, 3001, and 11287, respectively. The
incomplete records over the three years range from 299 to 721. This may be due to the fact that some students did not take the test at all. Their records were missing for various reasons. For example, students may have moved out of the state during the period. Since the missing number was small for each year (ranging from 2 to 5 percent), only the valid scale scores were used to describe the trends in achievement and achievement gaps. The table below (Table 4.1.1) provides an overall picture of the means and the standard deviations of the scale scores across the three years. The vertical scaling technique applied in this state’s assessments made it possible to look at and compare the scale scores across years and across grades. The table shows that the average scale scores increased since the year 2002-03.

**Table 4.1.1**
The Means and the Standard Deviations of the Overall Scale Scores

<table>
<thead>
<tr>
<th>Year</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1: 2002-2003</td>
<td>529.9</td>
<td>73.9</td>
</tr>
<tr>
<td>Year 2: 2003-2004</td>
<td>548.7</td>
<td>70.3</td>
</tr>
<tr>
<td>Year 3: 2004-2005</td>
<td>571.7</td>
<td>62.2</td>
</tr>
</tbody>
</table>

The table above provides an encouraging picture of student achievement for not only did the mean scale score increase across the three years, but the standard deviation of the scores decreased. The notion of standard deviation, by referring to the average distance between the scores and their mean score, describes how spread-out the scores were. A smaller value of standard deviation indicates that the variation of student scale scores decreased, and that students were becoming more homogeneous with respect to their math achievement. On one hand, the increase of mean scale scores suggests an
improvement of student academic performance, implying the progress of education quality; on the other hand, the decrease of variation implies that the differences between the students who performed well and those who performed poorly were narrowing, implying the reduction of education inequality. Therefore, the overall results appeared desirable since students seemed to have improved their academic performance together with their peers. The interpretation, however, should be cautious since a few studies suggested that the change of standard deviation might be related with the selection of testing models. For example, when the Item Response Testing model is applied, which is the case in this study, standard deviations of scale scores tend to decrease across years.

The general trend, without taking into account student demographic information and school/district characteristics, was not very revealing. Analyses were further disaggregated to provide a clue of whether the score gaps between different ethnic groups have decreased across the three years. Before answering this question, the trajectories of the scale scores were presented for each ethnic group (see Figure 4.1.1 and Figure 4.1.2).

![Figure 4.1.1](image1.png)

*Figure 4.1.1. The mean scales scores of the three ethnic groups.*

![Figure 4.1.2](image2.png)

*Figure 4.1.2. The standard deviations of the three ethnic groups.*
Figure 4.1.1 shows that the average scale scores of all the three ethnic groups improved significantly from year 1 (2002-03) to year 3 (2004-05). However, for each year, White students on average performed better than Black students; and Black students on average performed slightly better than Hispanic students. The differences in the mean scale scores between White and the two minority groups of students were approximately constant across the years. Figure 4.1.2 shows that in general the standard deviations for all the three ethnic groups decreased. The trend in standard deviation here was desirable: when the mean scale scores increased, the standard deviations of the scores decreased, and thus, the scale scores of these groups of students became more homogeneous. For example, Hispanic students had the lowest average scale scores, and the variance of their scores decreased most.

Several trends in the score gaps are revealed from the two figures above: (1) In general the achievement gaps between non-minority and minority students decreased. (2) The score gaps between White students and minority students were larger than the gaps between the two minority groups themselves. (3) The gaps between White and Hispanic students (H-W) were larger than the gaps between Black and White students (B-W).

In order to better illustrate the trends in the score gaps, Figure 4.1.3 does not only depict the lines of the score gaps between different ethnic groups, but also demonstrates the magnitude of the changes in the gaps between years. For example, the score gap between Hispanic and White students in the year 2002-03 was 57.4, meaning that in this year White students, on average, scored 57.4 points higher than Hispanic students. From
the year 2002-03 to 2003-04, the H-W gap decreased from 57.4 to 50.5 points. Therefore, the magnitude of the decrease is 6.9 points.

**Figure 4.1.3.** The Score Gaps between Ethnic Groups.

**Figure 4.1.4.** The Effect Sizes of the Score Gaps between Ethnic Groups.

Figure 4.1.3 shows that both gaps decreased significantly from year 2002-03 to 2004-05 while the decrease of the H-W gap seemed more dramatic than the B-W one. For example, from year 1 to year 2, the B-W gap did not decrease much. Meanwhile, the decreases of the H-W gap seemed consistent, while the B-W gap was not. A positive value of the magnitude of the change represents a decrease of the score gap (i.e. 6.9 in the figure) while a negative value indicates an increase of the gap. There is no negative value here, indicating that both gaps did not ever increase across the three years.

Although the above analysis has demonstrated the score differences for each year and the magnitude of the changes in the differences between years, it does not tell about the size of the difference and whether the difference is meaningful. Although significant differences were found between the average scores of minority and non-minority
students, it may be due to the large sample size in this study. Effect size is applied to take into account the association between sample size and likelihood of achieving statistical significance (see Figure 4.1.4). It aims to estimate the magnitude of the difference between two groups. The word “effect” here does not refer to the measure of treatment effect. Instead, it only indicates the standardized score difference between two ethnic groups. To calculate effect size, the formula of Cohen's $d$ is applied, using the means as well as the standard deviations of any two groups (see Formula 4.1).

Cohen's $d = \frac{M_1 - M_2}{\sigma_{pooled}}$

\[\sigma_{pooled} = \sqrt{\frac{(\sigma_{group1}^2 + \sigma_{group2}^2)}{2}}\]  (Formula 4.1)

Figure 4.1.4 shows the magnitudes of the effect sizes between White and minority students over the three years. For example, the effect size of 0.82 means that the mean score of Black students is nearly one standard deviation below the mean score of White students. This is substantial difference between White and Hispanic students. The effect sizes between White and Black students are also substantial ranging around 0.57. The figure is consistent with the previous results. For example, the effect size of H-W gap was bigger than the one of B-W gap, showing that the score differences between White and Hispanic students were larger than the differences between White and Black students. Meanwhile, the score differences between Hispanic and White students narrowed from 0.82 to 0.75. However, the figure shows that there was not much change in the gap between White and Black students after taking into account the standard deviation of the scores. Stated differently, the standardized difference between the scores of Black and White students decreased very slightly.
Gain Score

In addition to describing the overall pattern of the changes in scale scores over time as well as comparing the scores of different ethnic groups from one time period to another, the study aims to examine achievement and achievement gaps from the perspective of growth. Growth in the descriptive analyses refers to the gain score—the score gap between the starting year (year 2002-03) and the ending year (year 2004-05). Achievement gap then refers to the gain score gap instead of the scale score gap in the previous section.

ANOVA (Analysis of Variance)

Gain scores were compared among the ethnic groups. It was found that the scores of the three groups significantly differed from each other—Hispanic students gained most (Mean = 52.2), while white students gained least (Mean = 38.3). Meanwhile, the gain scores of Hispanic students had the largest variance (SD = 38.6) when the gain scores of White students were more homogeneous (SD = 34.4). The table below provides the description of gains scores at the group level.

Table 4.1.2
The Description of the Gain Scores of the Three Ethnic Groups

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>N</th>
<th>Mean Starting Score</th>
<th>Mean Gain Score</th>
<th>Std. Deviation</th>
<th>% of Not Gain</th>
<th>% of Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>534</td>
<td>504.7</td>
<td>45.9</td>
<td>35.2</td>
<td>6.8</td>
<td>93.3</td>
</tr>
<tr>
<td>Hispanic</td>
<td>2672</td>
<td>486.9</td>
<td>52.2</td>
<td>38.6</td>
<td>6.7</td>
<td>93.3</td>
</tr>
<tr>
<td>White</td>
<td>10321</td>
<td>545.0</td>
<td>38.3</td>
<td>34.4</td>
<td>11.8</td>
<td>88.1</td>
</tr>
<tr>
<td>Total</td>
<td>13527</td>
<td>529.9</td>
<td>41.34</td>
<td>35.8</td>
<td>10.7</td>
<td>89.3</td>
</tr>
</tbody>
</table>

N: Number of students;  Std Deviation: Standard deviation
Mean Gain Score: average gain score for one ethnic group
% of Not Gain: percentage of students whose scores did not increase between year 3 and year 1
% of Gain: percentage of students whose scores increased between year 3 and year 1
It was found that most of students’ scores increased across the three academic years (89.3%), while about 11% of students’ scores either did not increase or even decreased from the starting year (2002-03) to the ending year (year 2004-05). More starting scores of White students (11.8%), compared with the ones of Black (6.8%) and Hispanic students (6.7%), did not increase or even decreased.

The conclusion that minority students (Black and Hispanic) gained more along the three years cannot be simply drawn because: First, the sample sizes for non-White students were small ($N_{\text{Black}} = 534$, $N_{\text{Hispanic}} = 2672$) compared with their white peers ($N_{\text{White}} = 10321$). The assumption of homogeneity of variances was found violated when means were compared. When the variances of the three ethnic groups significantly differed from each other, the comparison of the means became invalid since the mean differences can be confounded with the variance difference. Second, some other important factors might contribute to the ethnic differences in gain scores. In this study, an important assumption was made that the test performance is reported on an interval scale so that a difference between a score of 250 and a score of 260 would represent the same difference in student math capability as would a difference between a score of 550 and a score of 560. Under this assumption, the gain score is not expected to be correlated with starting score, meaning where a student starts cannot predict how much a student gains. The table above shows, however, that the lower the average starting score a group had, the higher the mean gain score it contained. Therefore, starting score was included as a covariate in the analysis below to examine the assumption that there is no significant correlation between starting scores and gain scores.
Analysis of Covariance (ANCOVA)

ANCOVA was applied by including starting score as a covariate so that the correlation between gain score and starting score can be investigated. This approach aims to answer the question—did the gain scores differ among ethnic groups after adjusting for the starting scores? The results (See Table 4.1.4, Model 2, pg 78) showed that both ethnicity and starting score significantly contributed to the variation in gain score, the dependent variable. Hence, after adjusting for the variation in the starting scores, the gain scores still differed significantly among the ethnic groups.

This ANCOVA approach was based on a very important assumption that the covariate “starting score” was not correlated with the grouping factor “ethnicity”. To examine this assumption, the interaction between ethnicity and starting score was included in the ANCOVA model (See Table 4.1.4, Model 3, pg 78). The results showed that after including the interaction effect, ethnicity was not a significant predictor any more. The interaction effect indicated that the relationship between starting score and gain score was not constant across ethnic groups. Therefore, after adjusting for starting score as well as the interaction between starting score and ethnicity, ethnic groups did not differ in their mean gain scores any more.

To further disaggregate the analysis, the variable performance level was included to investigate whether students starting at different performance levels gained in different patterns (see Table 4.1.4, Model 4). For example, did students starting at the low performance level grow differently from the students starting at the high performance level? The scale scores in this study were assigned to four performance levels.
Performance level 3 indicates that students meet the state curriculum standards. The table below presents how the four performance levels at the starting year distributed by ethnicity.

Table 4.1.3
Distribution of Performance Levels at the Starting Year by Ethnicity

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>128 (22.4%)</td>
<td>210 (36.7%)</td>
<td>162 (28.4%)</td>
<td>70 (12.3%)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>857 (30.2%)</td>
<td>1111 (39.0%)</td>
<td>689 (24.3%)</td>
<td>173 (6.5%)</td>
</tr>
<tr>
<td>White</td>
<td>995 (9.3%)</td>
<td>2885 (26.7%)</td>
<td>4141 (38.3%)</td>
<td>2766 (25.6%)</td>
</tr>
<tr>
<td>Total</td>
<td>1980 (14.0%)</td>
<td>4206 (29.6%)</td>
<td>4992 (35.1%)</td>
<td>3009 (21.2%)</td>
</tr>
</tbody>
</table>

The results showed that students were not evenly distributed by performance level and ethnicity. More minority students were found to be at the lower levels of performance at the starting year. For example, 63.9% (38.3%+25.6%) of White students reached the level of being proficient (Level 3) or above (Level 4), while only 40.7% (28.4%+12.3%) of Black students and 30.8% (24.3%+6.5%) of Hispanic students reached these two levels.

To gain an understanding of how the gain scores changed over years, the covariates starting score and performance level, together with the grouping variable ethnicity, were included into the final model, where the interaction effects between and among these predictors were also considered (see Table 4.1.4, Model 4). The results showed that after including the covariate performance level and its interaction with ethnicity and starting score, all the predictors, as well as the interactions between and
among them, became significant. The changes from Model 3 to Model 4 suggest that students starting at different performance levels grew in different patterns.

The following table compares the results obtained from various approaches which are used to model the gain score gap.

Table 4.1.4

<table>
<thead>
<tr>
<th>Models</th>
<th>Approach</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>ANOVA</td>
<td>Ethnicity **</td>
</tr>
<tr>
<td></td>
<td>Grouping factor: Ethnicity</td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>ANCOVA</td>
<td>Ethnicity **</td>
</tr>
<tr>
<td></td>
<td>Grouping factor: Ethnicity</td>
<td>Starting score**</td>
</tr>
<tr>
<td></td>
<td>Covariate: Starting score</td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td>ANCOVA (with interaction effect)</td>
<td>Ethnicity **</td>
</tr>
<tr>
<td></td>
<td>Grouping factor: Ethnicity</td>
<td>Starting score**</td>
</tr>
<tr>
<td></td>
<td>Covariate: Starting score</td>
<td>Ethnicity x Starting score**</td>
</tr>
<tr>
<td>Model 4</td>
<td>ANCOVA (with interaction effect)</td>
<td>Ethnicity**</td>
</tr>
<tr>
<td></td>
<td>Grouping factor: Ethnicity</td>
<td>Starting score**</td>
</tr>
<tr>
<td></td>
<td>Covariate:</td>
<td>Performance level**</td>
</tr>
<tr>
<td></td>
<td>• Starting score</td>
<td>Ethnicity x Starting score**</td>
</tr>
<tr>
<td></td>
<td>• Performance level</td>
<td>Ethnicity x Performance level**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Starting score x Performance level**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ethnicity x Starting score x Performance level **</td>
</tr>
</tbody>
</table>

*p<.05  **p<.01

To explain the significant interaction effects in Model 4, the ANCOVA analysis in Model 3 was conducted, respectively, for the students starting at each performance level. It was found that for students starting at the lowest performance level, not only did their mean gain scores differ by ethnicity, but their starting scale scores were also a significant predictor of their gain scores. The interaction effect between ethnicity and starting score was found significant too, indicating that the prediction of the starting
scores in the gain scores differed among ethnic groups. However, for the other three performance levels, no predictor as well as interaction effect was found significant.

Figure 4.1.5 shows a negative relationship between mean gain scores and performance levels—the mean gain scores significantly decreased when performance level increased from level 1 to level 4. Interaction is said to exist when the lines corresponding to the different ethnic groups cross. However, only at the performance level 1, gain scores differed significantly among ethnic groups—Hispanic students gained most while White students gained least. At the other three levels, no significant difference was found among the mean gain scores of ethnic groups.

![Figure 4.1.5. Mean Gain Scores across Performance Levels.](image)

**Summary**

Both scale scores and gain scores were employed to describe the general trends in the achievement and the achievement gaps. There were scale score gaps between
minority and non-minority students over the three academic years. The Hispanic-White score gaps narrowed while the Black-White gaps only decreased slightly. The application of effect size, however, presented a different picture. The standardized differences in the scores between Hispanic and White students narrowed between year one and year two but remained almost constant between year two and year three. The standardized differences in the scores between Black and White students were first widened slightly and then remained constant.

The notion of gain score was applied to capture the trends in achievement growth and growth gaps. Minority students, on average, were found to start with lower scores but end with higher gain scores. The analyses of covariance were then applied to investigate whether gain scores differed significantly by ethnicity after adjusting for a few covariates as well as their interactions. The covariate starting score was found to be significantly correlated with ethnicity. However, after adjusting for the interaction between ethnicity and starting score, the result showed that gain scores did not differ significantly by ethnicity. The analyses were then disaggregated into different performance levels. For example, for students starting at performance level 1 (the lowest level), their mean gain scores differed significantly by ethnicity even after adjusting for the covariates starting score and its interaction with ethnicity. However, for students starting at the other three performance levels, the gain scores were not found to differ by ethnicity. Therefore, for students starting at different levels, the growth patterns may be different. Ethnic differences in gain scores seemingly only existed in the group of students who started at a very low performance level.
District Level

In addition to the general trends in achievement and achievement gap, of which the calculation was based on the achievement scores of all students, here arises the question of whether the average achievement scores and the average achievement gaps of schools and school districts differ from each other. The graph below (Figure 4.1.6) demonstrates the distribution of average scale scores at the ending year (2004-05) for different ethnic groups in 14 districts. Each dot represents an average scale score of an ethnic group in one district. For example, for District No.12, the mean scale scores were 605.0, 564.8, and 581.5 for Black, Hispanic and White students respectively. The average scale scores for the 14 districts range between 520 and 620 and largely cluster together based on student ethnicity. The mean scores of Black students, however, spread out to a great extent. For example, the mean score of Black students in District No.12 is far above the scores of any other groups in the districts. It was found that the cohort in this district only contained 4 (1.2%) Black students and they all scored high (ranging from 560 to 620).

The results demonstrated by Figure 4.1.6 were consistent with the previous conclusion that White students, on average, performed better than Black and Hispanic students in the mathematics state assessments. In every district except District No. 12, White students scored higher than their Hispanic and Black peers. However, the average scores differed district by district and overlaps existed among ethnic groups. For example, the average score of White students in District No. 3 was found to be lower than the average scores of Black and Hispanic students in a few other districts. The variance
among districts confirms the necessity of including the district level into the Hierarchical Linear Model for further analysis.

Note that only 13 districts contained Black students. District No. 9 did not contain one single Black student in the cohort. District No. 14 only contained one black student in the cohort and this student scored very low (470.0). The outlier was not included in the figure.

<table>
<thead>
<tr>
<th>District Number</th>
<th>Average Scale Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>620</td>
</tr>
<tr>
<td>2</td>
<td>600</td>
</tr>
<tr>
<td>3</td>
<td>580</td>
</tr>
<tr>
<td>4</td>
<td>560</td>
</tr>
<tr>
<td>5</td>
<td>540</td>
</tr>
<tr>
<td>6</td>
<td>520</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
</tr>
</tbody>
</table>

*Figure 4.1.6. Average Scale Scores of Each District for Different Ethnic Groups.*

The graph below (Figure 4.1.7) illustrates the distributions of average gain scores for different ethnic groups in 14 districts. Each dot represents an average gain score for an ethnic group in one district. For example, for District No.12, the average gain scores were 39.8, 32.5, and 29.5 for Black, Hispanic and White students respectively. The mean gain scores for the 14 districts range between 20 and 80 and largely cluster together based on student ethnicity. The mean gain scores of Black students, however, spread out to a
great extent. The results were consistent with the previous conclusion that the gain scores of Hispanic students, on average, were higher than the gain scores of the other two ethnic groups. In most districts Hispanic students gained most while in four districts, Black students gained most. White students gained least in almost every district except in District No.10 where the average gain scores for White students and Black students were very close. Note that District No. 14 only contained one black student in the cohort and this student scored very low (470.0) and gained very little (13.0). The outlier was not included in the figure.

Figure 4.1.7. Average Gain Scores of Each District for Different Ethnic Groups over Two Years.

After describing the average scale scores at the district level, we now demonstrate the achievement gaps between White and minority students at the district level. Figure 4.1.8 illustrates the distributions of the achievement gaps between White and Hispanic as well as between White and Black students in the 14 districts. Each dot represents the
difference in the average scores between two different ethnic groups in one district. For example, for District No.12, the gap between the average scores of White students and Black students was -23.5, indicating that White students in this district scored 23.5 points lower than their Black peers; and the gap between White and Hispanic students was 16.7, indicating that White students in this district scored 16.7 points higher than their Hispanic peers. This district is an outlier since it is the only one where White students scored lower than their Black peers (the dot is below the zero line when the value of the gap is negative). It was found that there were only four Black students in this district and they all scored high. Except for this district, the average scores of White students in all other districts were found to be higher than the score of their minority peers.

Figure 4.1.8. Achievement Gaps between Black and White Students and between Hispanic Students at the District Level.

---

6 This district is not included in the HLM analysis of the Black-White achievement growth gaps due to the very limited sample size of Black students
The graph further shows that the overlaps exist between the two gaps - Hispanic-White (H-W) gap and Black-White (B-W) gap, although the latter one is more spread out.

Figure 4.1.9 illustrates the distributions of the gain score gaps between White and Hispanic as well as between White and Black students in the 14 districts. Each dot represents the difference in the average gain scores between two different ethnic groups in one district. For example, for District No.12, the gap between the average gain scores of White students and Black students was -10.3, indicating that White student in this district gained 10.3 points fewer than their Black peers; and the gap between White and Hispanic students was -3.1, indicating that White student in this district gained 3.1 points fewer than their Hispanic peers. In almost every district except District No.10, the average gain scores of White students were found to be lower than the ones of their minority peers (the dots are below zero line when the value of the gap is negative). In District No.10, the average gain score of White students was almost the same as the one of their Black peers (the dot is close to zero when the gap is close to zero).

The graph further shows that the overlaps exist between the two gaps—Hispanic-White (H-W) gap and Black-White (B-W) gap, although the latter one is more spread out. For example, for District No.11, the gain score of White students is 43.9 points higher than the gain score of their Black peers.
Figure 4.1.9. Gain Score Gaps between Black and White Students and between Hispanic Students at the District Level.

**School Level**

In addition to the average scale scores and the average achievement gaps at the district level, the average scores\(^7\) and the score gaps at the school level were examined. Figure 4.1.10 depicts the distributions of scale score gaps at the school level between Hispanic and White students (H-W gap) and between Black and White students (B-W gap). For the H-W gap (X-axis), most of the schools cluster between 0 and 100 with only several schools out of this range. It indicates that in most of the schools, the average scale scores of White students were higher than the scores of Hispanic students. The dots on the left of the zero line, with a negative value, indicate that Hispanic students on average outperformed their White peers in the two schools (School No. 1 & School No.30). For example, in School No.30, Hispanic students on average scored 35.5 points higher. For

---

\(^7\) The description of the scores at the school level presents in the next section “comparisons among levels.”
the B-W gap (Y-axis), most of the schools cluster between -50 and 150. The range is much wider than the H-W gap. The dots below the zero line indicate that Black students, on average, performed better than their White peers in their schools. There are a lot more dots above the zero line, indicating that in these schools Black students had lower average scores than their White peers. For example, for School No. 92 and School No. 67, Black students on average scored about 170 points lower than their White peers.

Most of the dots cluster in Quadrant One, suggesting that in these schools, both Black and Hispanic students scored lower than their White peers. Only one school was found in Quadrant Two, suggesting that there was only one school where both Hispanic and Black students scored higher than their White peers.

![Graph showing the average H-W and B-W scale score gaps in every school.](image)

**Figure 4.1.10.** The Average H-W and B-W Scale Score Gaps in Every School.

The analysis at the district level has showed that the scale score gap was different from the gain score gap. Now the two gaps at the school level are included in the same
graph to further illustrate the difference. Figure 4.1.11 demonstrates the distributions of the gaps between Hispanic and White students in every school. For the scale score gap (X-axis), the result was consistent with Figure 4.1.10 when most of the dots cluster between 0 and 100, suggesting the average scale scores of White students in the schools were higher than the scores of Hispanic students. For the gain score gap (Y-axis), most of the schools cluster between -50 and 25 with only several schools out of this range. A negative value here indicates White students gained less than Hispanic students (see the dots below the zero line).

![Figure 4.1.11. The Average H-W Scale Score and Gain Score Gaps for Different Schools.](image)

Most of the schools cluster in Quadrant Three, where Hispanic students scored lower but gained more than their White peers in the same schools. In some other schools
(Quadrant Two), Hispanic students scored lower and gained less. There were 5 schools where Hispanic students scored higher and gained more (Quadrant Four). The schools in Quadrant Two and Quadrant Four show that students had different growth patterns when the association between scale score and gain score is considered.

![Figure 4.1.12](image.png)

**Figure 4.1.12.** The Average B-W Scale Score and Gain Score Gaps for Different Schools.

The gap between Black and White students at the school level is another story. Compared with Figure 4.1.11, Figure 4.1.12 shows (1) that there were more schools where Black students, on average, scored higher than their White peers (see the dots on the left side of the zero line of X-axis), and (2) that there were more schools where Black students gained less than their White peers (see the dots above the zero line of Y-axis). There were several schools where Black students scored higher but gained less than their White peers (see Quadrant One). There was no such school for Hispanic students in
Figure 4.1.10. Instead of mainly clustering in Quadrant Three (scored lower but gained more) as Hispanic students, Black students more evenly distributed in the four quadrants. This suggests that not only is the variance of the scores Black students larger, but their growth patterns are more heterogeneous.

**Comparisons among Levels**

In order to probe under the surface of the general trend in the overall scores and to give a comprehensive view of achievement and achievement gaps, the scale scores and the gain scores were compared at three levels—district, school, and overall (individual). Table 4.1.5 demonstrates what the means of the scale scores at the three levels were and how these scores spread out.

For Hispanic students, the overall mean based on the scores of all the students (537.5) was slower than the mean of the average scores at the school level (544.9) and the district level (541.8). This suggests that Hispanic students who performed relatively well tended to go to small schools and districts. Stated differently, Hispanic students in large schools/districts performed more poorly than those in small schools/districts.

The story for White students was different. The overall mean of the scale scores (582.0) was slightly higher than the mean of the average scores at the school level (577.9) and the district level (578.1). The difference suggests that White students who performed relatively well tended to go to large schools and districts. However, since the mean scores at the three levels were very close, it implied that White students distributed almost evenly across schools and districts with respect to their math scale scores.
For Black students, the story was not very straightforward when the overall mean was larger than the mean at the district level but smaller than the mean at the school level. It could suggest that Black students who performed relatively well tended to go to large districts but small schools. The conclusion should be taken cautiously since the distribution of the scores of Black students was irregular: (1) the sample size of Black students was much smaller than the ones of the other two groups at all three levels; and (2) the range of the scores for Black students was a lot wider than the ones for the other two groups. The table below confirms that the scores of Black students at all three levels scatter more widely when their standard deviations are much larger than the ones of the other two groups.

Table 4.1.5
Comparing the Average Scale Scores at Three Levels - District, School and Overall

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>District</th>
<th>School</th>
<th>Overall</th>
<th>District</th>
<th>School</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>540.9</td>
<td>553.1</td>
<td>547.2</td>
<td>44.9</td>
<td>47.5</td>
<td>62.8</td>
</tr>
<tr>
<td>Hispanic</td>
<td>541.8</td>
<td>544.9</td>
<td>537.5</td>
<td>13.2</td>
<td>28.5</td>
<td>60.5</td>
</tr>
<tr>
<td>White</td>
<td>578.1</td>
<td>577.9</td>
<td>582.0</td>
<td>13.5</td>
<td>27.0</td>
<td>58.9</td>
</tr>
</tbody>
</table>

To better understand student growth at all three levels, Table 4.1.6 displays the means and standard deviations of the gain scores. The table shows that for Hispanic and White students, the means at different levels were very close, suggesting that these students may distribute evenly across schools and districts with respect to their gain scores. Stated differently, there was no distinct association between school/district size and the gain scores. The distribution of the gain scores of Black students, however, was different from these two groups. The overall mean of the gain scores (45.9) was smaller.
than the ones at the other two levels (50.3 & 50.5), suggesting Black students who gained more tended to go to small districts and schools. Interpreted differently, Black students in small districts on average scored higher than those in big districts. Consistent with Table 4.1.5, Table 4.1.6 shows that the gain scores of Black students scatter more widely at all three levels.

Table 4.1.6
Comparing the Average Gain Scores at Three Levels - District, School and Overall

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>District</td>
<td>School</td>
</tr>
<tr>
<td>Black</td>
<td>50.3</td>
<td>50.5</td>
</tr>
<tr>
<td>Hispanic</td>
<td>52.5</td>
<td>50.0</td>
</tr>
<tr>
<td>White</td>
<td>42.2</td>
<td>40.9</td>
</tr>
</tbody>
</table>

In order to obtain a deeper understanding of achievement gaps, the scale score gaps and the gain score gaps are disaggregated into the district and school levels. Table 4.1.7 describes the distributions of the scale score gaps at these two levels. For example, the mean of the average gaps between Black and White students (B-W gap) in all schools was 24.6. It means that on average Black students scored 24.6 points lower than their White peers in the same schools. As to the district level, the mean of the B-W gaps was 36.1, meaning on average Black students scored 36.1 points lower than their White peers in the same districts. The overall mean B-W gap (34.2) is close to the mean at the district level (36.1), but wider than the one at the school level (24.6). This suggests that the students, no matter whether they were White or Black students, tended to score more closely if they were in the same schools. The H-W gap reveals a similar pattern when the gap at the overall level (44.0) is wider than the mean gap at the school level (32.3) and
the district level (36.4). It suggests that students, no matter whether they were White or Hispanic, tended to score closely if they were in the same districts or schools.

In addition, the variances of the gaps at the district and school levels confirm the necessity of examining the random effects of schools and districts since they differ from each other with regard to their scale score gaps.

Table 4.1.7

<table>
<thead>
<tr>
<th>Gap</th>
<th>School Level</th>
<th>District Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>B-W</td>
<td>73</td>
<td>24.6</td>
</tr>
<tr>
<td>H-W</td>
<td>94</td>
<td>32.3</td>
</tr>
</tbody>
</table>

In addition to the scale score gaps, Table 4.1.8 describes and compares the gain score gaps at the school and district levels. All the values in this table are negative, indicating that White students on average gained less across schools and districts. For example, for the B-W gap, the mean of the gain score gaps at the school level was -10.8, meaning that White students on average gained 10.8 points fewer than Black students across all the schools.

The differences of the mean gain score gaps among the overall, school and district levels seemed small for both the B-W gap (-12.4, -10.8 & -8.4) and the H-W gap (-8.5, -9.2 & -10.3). It suggests that the gain score gaps may distribute in a similar pattern at the three levels. Stated differently, students may be distributed relatively evenly across schools and districts with regard to their gain score. This also suggests there seemed no association between the gain scores and school/district size.

97
Table 4.1.8
Comparing the Gain Score Gaps at the School Level and the District Level

<table>
<thead>
<tr>
<th>Gap</th>
<th>School Level</th>
<th>District Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>B-W</td>
<td>71</td>
<td>-10.8</td>
</tr>
<tr>
<td>H-W</td>
<td>94</td>
<td>-9.2</td>
</tr>
</tbody>
</table>

Summary

The average scale scores and the average gain scores were calculated to understand the trends in achievement and achievement gaps at the school and district levels. At both levels, the findings were consistent with the results obtained from the general trend based on individual scores (overall level). For example, Hispanic students, on average, scored lowest but gained most at all three levels. Nevertheless, there was considerable variability among districts and schools; overlaps were found among ethnic groups in different schools and districts; and outliers were identified when the distributions of achievement and achievement gaps at some schools/districts were not consistent with the general trend. These inconsistencies further confirmed the necessity of including the school and district levels in multi-level analysis which takes into consideration the variability among schools and districts.

The scale scores and the gain scores were compared at the three levels—district, school, and overall. It was found that Hispanic students who performed relatively well tended to go to small schools and districts while White students seemed evenly distributed across schools and districts with respect to their math scale scores. The distribution of the scale scores of Black students was somehow irregular and the story was not very straightforward. As to the gain scores, Black students who gained more
tended to go to small districts and schools while for White and Hispanic students there
seemed no distinct association between their gain scores and school/district size.
However, considering that the sample size of Black students is very small, the
observation should be taken cautiously since it can be subjective to the random
fluctuations of individual scores of Black students.

The comparisons with regard to the scale score gaps showed that no matter what
ethnic group students belonged to, they tended to score closely if they were in same
schools and districts. The comparison of the gain score gaps revealed a different story.
Students distributed mostly evenly across schools and districts with respect to their gain
scores. The HLM models are to be applied to examine the random variation in the
achievement gaps and achievement growth gaps in order to help understand whether the
patterns of student achievement growth varied across schools and districts.

**Independent Variables**

Rather than only identifying whether there were achievement gaps, we also
investigated what school and school district characteristics can account for the variation
in achievement and achievement gaps. The school and district characteristics include
school/district size, teacher-student ratio, minority percentage, percentage of the students
who received free or reduced-price lunch, percentage of Special Ed/IEP Students, and
percentage of English Language Learner Students (ELL), etc. Among all these
independent variables, the index of racial diversity at the district level and the mobility
index at the school level are especially interesting since this study is a longitudinal study with regard to racial differences in education achievement.

**Index of Racial Diversity**

As suggested in the chapter on research design, the index of racial diversity was employed as an indicator of the degree of segregation for the school districts that participated in this study. While the school segregation indices in the literature generally sum up the racial distributions of all the schools in a district, this study is only concerned with the cohort of grade 6 in 2002-03, grade 7 in 2003-04, and grade 8 in 2004-05. Hence, the formula of the H segregation index is applied to obtain a cohort index of racial diversity. The calculation of the cohort index involves 15,707 students, 97 schools, and 14 districts. The table below presents the detailed information of the index of racial diversity as well as the percentage of each racial group in the districts.

The index ranges from 0.01 to 0.12. The range is not great, suggesting that the districts are relatively homogeneous with respect to their relative degree of racial segregation. Moreover, all of the index values are close to 0 instead of 1, indicating that segregation was not serious in these districts (Mean=0.05, N=14). To be more specific, the schools in these districts had similar racial compositions. Stated differently, the individual schools, to a great extent, were as diverse as their district as a whole. In sum, this index is indicative of how evenly groups distribute among schools in a district. It shows that the schools in the suburban districts distribute evenly with regard to their proportions of different ethnic groups.
Table 4.1.9
The Index of Racial Diversity and the Distribution of Racial Groups in the Districts

<table>
<thead>
<tr>
<th>No.</th>
<th>Index</th>
<th>Ethnicity (number and percentage)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>American Indian</td>
<td>Asian</td>
</tr>
<tr>
<td>1</td>
<td>.06</td>
<td>24(9.0%)</td>
<td>130</td>
</tr>
<tr>
<td>2</td>
<td>.04</td>
<td>6(7.7%)</td>
<td>10(1.6%)</td>
</tr>
<tr>
<td>3</td>
<td>.03</td>
<td>5(6.6%)</td>
<td>67(10.1%)</td>
</tr>
<tr>
<td>4</td>
<td>.02</td>
<td>4(2.4%)</td>
<td>5(1.5%)</td>
</tr>
<tr>
<td>5</td>
<td>.06</td>
<td>17(5.5%)</td>
<td>200(6.3%)</td>
</tr>
<tr>
<td>6</td>
<td>.05</td>
<td>10(9.9%)</td>
<td>35(2.9%)</td>
</tr>
<tr>
<td>7</td>
<td>.12</td>
<td>11(9.9%)</td>
<td>48(3.1%)</td>
</tr>
<tr>
<td>8</td>
<td>.11</td>
<td>17(6.6%)</td>
<td>112(5.0%)</td>
</tr>
<tr>
<td>9</td>
<td>.05</td>
<td>2(6.6%)</td>
<td>4(9.6%)</td>
</tr>
<tr>
<td>10</td>
<td>.02</td>
<td>4(1.2%)</td>
<td>15(3.2%)</td>
</tr>
<tr>
<td>11</td>
<td>.09</td>
<td>0(0.0%)</td>
<td>4(9.6%)</td>
</tr>
<tr>
<td>12</td>
<td>.04</td>
<td>19(5.6%)</td>
<td>3(1.6%)</td>
</tr>
<tr>
<td>13</td>
<td>.03</td>
<td>14(1.1%)</td>
<td>13(9.3%)</td>
</tr>
<tr>
<td>14</td>
<td>.02</td>
<td>4(8.8%)</td>
<td>1(6.6%)</td>
</tr>
<tr>
<td>Missi</td>
<td>-</td>
<td>8(3.3%)</td>
<td>6(2.5%)</td>
</tr>
<tr>
<td>Total</td>
<td>-</td>
<td>145(9.9%)</td>
<td>653(4.2%)</td>
</tr>
</tbody>
</table>

The notion of minority percentage is introduced to the model. This measure is complementary to the index of racial diversity which is based on the specific proportion of each minority group in schools and school districts. The percentage of minority students in districts ranges from 10% to 60%, with the mean 28.6% (N=14). The percentage of minority students in schools ranges from 0% to 72%, with the mean percentage 25.1% (N=97). The variation of this measure is larger than the index of racial diversity. This is not surprising considering the measure does not take into account the information on how evenly groups are distributed among schools in a district. In sum, the indicators of ethnicity used in this study contain the information in different aspects and at different levels by including (1) individual ethnicity, (2) percentage of minority
students in a school, and (3) index of racial diversity in a district (H index). The table below demonstrates the distribution of the minority percentages in schools and districts.

Table 4.1.10

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Districts</td>
<td>14</td>
<td>10.0%</td>
<td>60.0%</td>
<td>28.6%</td>
<td>13.9%</td>
</tr>
<tr>
<td>Schools</td>
<td>97</td>
<td>0.0%</td>
<td>72.0%</td>
<td>25.1%</td>
<td>17.4%</td>
</tr>
</tbody>
</table>

**Index of Mobility**

Mobility needs to be taken into account in the longitudinal analysis so that the model can better capture what is going on in the schools system over years. Table 4.1.11 describes mobility at the student (individual) level between year 2002-03 (year 1) and year 2003-04 (year 2) as well as between year 2003-04 (year 2) and 2004-05 (year 3). It shows that most of the students did not move over the years. For example, 83.2% of students stayed in their schools between year 1 and year 2 and 90.0% of students stayed between year 2 and year 3.

Table 4.1.11

<table>
<thead>
<tr>
<th></th>
<th>Non-mover</th>
<th>Bureaucratic- Mover</th>
<th>Mover</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between year 1 &amp; 2</td>
<td>83.2%</td>
<td>6.8%</td>
<td>10.0%</td>
</tr>
<tr>
<td>Between year 2 &amp; 3</td>
<td>90.0%</td>
<td>1.1%</td>
<td>8.9%</td>
</tr>
</tbody>
</table>

Some students moved out of their schools because they reached the highest grade since some of the schools provide limited grade expansions. These students are called bureaucratic movers. For this cohort, the number of bureaucratic movers is small (6.8%
and 1.1%) since most of the schools containing grade 6 generally have the other two grades (grade 7 & 8). However, the percentages of the movers, who moved out of their schools for non-bureaucratic reasons, are noticeable (10.0% & 8.9%). Special attention is given to these students when the relationship between mobility and student achievement is investigated.

As to school mobility, it was found that on average 7.8% of the students either moved in or moved out of their schools. However, the range of school mobility is pretty broad, considering that in one school 56% of students moved and in another school no student moved at all. Note that only the students who moved for the non-bureaucratic reasons are taken into consideration to estimate the school mobility.

**Other Variables**

**District Level**

In addition to the index of racial diversity, there are some other independent variables at the district level, such as district size (total student number), percentage of Special Ed/IEP Students, and percentage of English Language Learner (ELL) Students. Table 4.1.12 describes the distributions of these variables. Note that all the numbers below are based on the overall information of the districts rather than the information of the cohort interested in this study. On average, a district had 10.8% of special-ed students and 12.7% of ELL students. The district size varied greatly from a few thousand students to almost fifty thousand students. The variation in pupil teacher ratio, however, was small, indicating districts did not differ much as to how many students shared a teacher.
Table 4.1.12
The Other Independent Variables at the District Level

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pupil-teacher ratio</td>
<td>14.3</td>
<td>20.0</td>
<td>16.7</td>
<td>2.4</td>
</tr>
<tr>
<td>Special-ed students (%)</td>
<td>8.0%</td>
<td>14.0%</td>
<td>10.8%</td>
<td>1.6%</td>
</tr>
<tr>
<td>ELL students (%)</td>
<td>2%</td>
<td>32%</td>
<td>12.7%</td>
<td>10.9%</td>
</tr>
<tr>
<td>Total student number</td>
<td>3883</td>
<td>47818</td>
<td>15964.7</td>
<td>13512.6</td>
</tr>
</tbody>
</table>

School Level

At the school level, in addition to the mobility index, other independent variables include school size (total student number), percentage of the students who received free or reduced-price lunch, and pupil-teacher ratio. Table 4.1.13 describes the distributions of these variables. All the school variables here are also based on the overall information of the schools instead of the information of the cohort of students.

Table 4.1.13
The other independent variables at the school level

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total student number</td>
<td>12.0</td>
<td>1982.0</td>
<td>636.2</td>
<td>365.1</td>
</tr>
<tr>
<td>Pupil teacher Ratio</td>
<td>2.8</td>
<td>32.0</td>
<td>16.3</td>
<td>3.4</td>
</tr>
<tr>
<td>FRL students (%)</td>
<td>0.0</td>
<td>100.0</td>
<td>27.0</td>
<td>21.0</td>
</tr>
</tbody>
</table>

On average, a school had about 636 students. The school size varied greatly from having only 12 students (No.3499) to having 1,982 students in the largest school. On average, 27.0% of the students in a school received free or reduced-price lunch. When there was a school where no students received free or reduced-price lunch, there was another school where every student did so. The socio-economic statuses of the schools varied greatly. Compared to the percentage of FRL students, the pupil teacher ratio had less variation when the standard deviation was relatively small. The average pupil teacher
ratio was 16.3, meaning on average, about every 16 students shared one teacher in a school.

**Correlation**

Correlational analyses were conducted to investigate the relationship between independent variables and the achievement and achievement gaps at different levels. The analyses include the investigations of the relationships (1) between student background characteristics and student achievement, (2) between the school/district characteristics and the mean achievement scores at the school and district levels, and (3) between the school/district characteristics and the achievement gaps at the school and district levels.

**Student Level**

Student background variables include ethnicity, gender, and student mobility. The relationship between ethnicity and achievement has been illustrated in the analyses of the achievement gaps between minority and non-minority students. The table below investigates the relationships between other student variables and the achievement scores.

<table>
<thead>
<tr>
<th>Table 4.1.14</th>
<th>Gender Differences in Student Achievement Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Scale Score</td>
</tr>
<tr>
<td></td>
<td>Number</td>
</tr>
<tr>
<td>Male</td>
<td>7262</td>
</tr>
<tr>
<td>Female</td>
<td>6984</td>
</tr>
</tbody>
</table>

Table 4.1.14 shows that there was a small gender difference in student achievement scores. Male students, on average, scored about 1 point higher but gained about 1 point fewer than female students.
Student mobility is another variable at the individual level which can be related with student achievement. Since each student had two chances to move along the three years, they are categorized as Non-mover (who stayed at their school over the three years), Bureaucratic Mover (who only moved once because they reached the highest grade of their schools), Mover-once (who moved once for non-bureaucratic reasons), Mover-twice-A (who moved twice and one of the moves was bureaucratic one), and Mover-twice-B (who moved twice for non-bureaucratic reasons).

Table 4.1.15 displays the number of movers for each category and their mean scale scores over three years. Non-movers were found to score the highest among all these groups, suggesting there was a negative correlation between mobility and student achievement.

<table>
<thead>
<tr>
<th>Group</th>
<th>Non-mover</th>
<th>Bureaucratic Mover</th>
<th>Mover (Once)</th>
<th>Mover (Twice)-A</th>
<th>Mover (Twice)-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Cases</td>
<td>10615</td>
<td>934</td>
<td>2000</td>
<td>81</td>
<td>236</td>
</tr>
<tr>
<td>Mean</td>
<td>556.4</td>
<td>540.7</td>
<td>524.5</td>
<td>514.1</td>
<td>510.1</td>
</tr>
</tbody>
</table>

1. Mover (Twice)-A: Students who moved twice, but once was due to the bureaucratic reason
2. Mover (Twice)-B: Students who moved twice, and both were due to the non-bureaucratic reasons

Table 4.1.16 shows the multiple comparisons of the average scale scores across three years between the groups of movers. The findings include: (1) the students who did not move scored significantly higher than any other groups; (2) the students who moved only because they reached the highest grade of their schools scored significantly higher than those who moved for the non-bureaucratic reasons; (3) the students who moved once scored significantly higher than those who moved twice; and (4) no significant difference
was found between the groups of the students who moved twice but due to different reasons (Mover A and Mover B).

Table 4.1.16  

*Multiple Comparisons Regarding the Difference in the Average Scores*

<table>
<thead>
<tr>
<th>Group</th>
<th>Non-mover</th>
<th>Bureaucratic Mover</th>
<th>Mover (Once)</th>
<th>Mover (Twice)-A</th>
<th>Mover (Twice)-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-mover</td>
<td>-</td>
<td>15.7**</td>
<td>31.9**</td>
<td>42.2**</td>
<td>46.3**</td>
</tr>
<tr>
<td>Bureaucratic Mover</td>
<td>-</td>
<td>-</td>
<td>16.1**</td>
<td>26.5**</td>
<td>30.6**</td>
</tr>
<tr>
<td>Mover (Once)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>10.4</td>
<td>14.4</td>
</tr>
<tr>
<td>Mover (Twice)-A</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4.0</td>
</tr>
<tr>
<td>Mover (Twice)-B</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

1. Mover (Twice)-A: Students who moved twice, but once was due to the bureaucratic reason
2. Mover (Twice)-B: Students who moved twice, and both were due to the non-bureaucratic reasons
3. *p<.05  **p<.01 (without family wise error adjustment)

**School Level**

Table 4.1.17 presents the relationships between independent variables and the average achievement scores as well as the achievement gaps at the school level. All the independent variables were found to be correlated with the gain scores at the school level except for school mobility rate. Therefore, although both student mobility and school mobility were found negatively correlated with the achievement schools, the school mobility rate was not a significant predictor of the average gain scores at schools. As to the correlation of gain score, the results showed that (1) the higher percentage of minority students in a school, the higher gain score was for the school; (2) the fewer students were in a school, the higher the average gain score was for the school; (3) the lower pupil-teacher ratio was in a school, the higher the average gain score was for the school; (4) the more students who received free/reduced-price lunch in a school, the higher the gain score was for the school. The results were understandable since the previous analyses
have found that minority students gained more than their White peers in schools, and that these minority students tended to go to the schools with lower social economic status.

A significant correlation was also found between average achievement scores and percentage of the students who received free or reduced-price lunch. The more students received free or reduced-price lunch in a school, the lower the average achievement score the school had. No school-level variable was found to significantly correlate with the scale score gaps and gain score gaps.

### Table 4.1.17
Correlations between the Achievement/Achievement Gaps and the Independent Variables at the School Level

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Achievement</th>
<th>Gain Score</th>
<th>Scale Gap (B-W)</th>
<th>Scale Gap (H-W)</th>
<th>Gain Gap (B-W)</th>
<th>Gain Gap (H-W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minority percentage</td>
<td>-.67**</td>
<td>.42**</td>
<td>-.08</td>
<td>-.16</td>
<td>-.16</td>
<td>-.04</td>
</tr>
<tr>
<td>School mobility rate</td>
<td>-.23*</td>
<td>.13</td>
<td>-.19</td>
<td>.032</td>
<td>.228</td>
<td>.107</td>
</tr>
<tr>
<td>Total student number</td>
<td>.09</td>
<td>-.28**</td>
<td>.03</td>
<td>.04</td>
<td>.16</td>
<td>.11</td>
</tr>
<tr>
<td>Pupil teacher Ratio</td>
<td>.13</td>
<td>-.29**</td>
<td>-.04</td>
<td>-.09</td>
<td>.30</td>
<td>.06</td>
</tr>
<tr>
<td>FRL students (%)</td>
<td>-.59**</td>
<td>.34**</td>
<td>-.11</td>
<td>.03</td>
<td>-.5</td>
<td>.16</td>
</tr>
</tbody>
</table>

*p<.05   ** p<.01 (without family wise error adjustment)

### District Level

Table 4.1.18 presents the relationships between independent variables and the average achievement as well as the achievement gaps at the district level. Among the five district variables, the index of segregation (H) was the only one significantly correlated with the achievement and the achievement gaps. Significant correlations were found (1) between the index of segregation and the scale score gap between Hispanic and White students, and (2) between the index of segregation and the gain score gap between Hispanic and White students. The first positive correlation indicates that the more
segregated a district was, the wider the H-W achievement gap in the district was. The second negative correlation implies that the more segregated a district was, the narrower the H-W gain score gap was. Note that a segregated district does not necessarily mean the district was not diverse. It was more of an indication of within-district homogeneity. The more schools’ racial compositions were different from the average composition of a district, the more segregated the district was.

No other district characteristics were found to be significantly correlated with the district average achievement and the district average achievement gaps.

Table 4.1.18
Correlations between the Achievement and Achievement Gap and the Independent Variables at the District Level

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>H (segregation index)</td>
<td>.24</td>
<td>.31</td>
<td>-.08</td>
<td>.67**</td>
<td>-.43</td>
</tr>
<tr>
<td>Pupil-teacher Ratio</td>
<td>.05</td>
<td>.07</td>
<td>.03</td>
<td>.25</td>
<td>-.23</td>
</tr>
<tr>
<td>Special-ed student (%)</td>
<td>.14</td>
<td>-.10</td>
<td>-.03</td>
<td>.19</td>
<td>-.11</td>
</tr>
<tr>
<td>ELL student (%)</td>
<td>.06</td>
<td>.09</td>
<td>-.10</td>
<td>.36</td>
<td>-.11</td>
</tr>
<tr>
<td>Total student number</td>
<td>.13</td>
<td>-.10</td>
<td>-.04</td>
<td>.20</td>
<td>-.11</td>
</tr>
</tbody>
</table>

*p<.05  **p<.01 (without family wise error adjustment)

Summary

The analyses of this section provided a picture of the achievement scores of different ethnic groups were and how they were distributed. Moreover, the achievement gaps were displayed at the overall (individual), school and district levels, respectively. All theses analyses, however, were not comprehensive and thorough enough since the connections among the levels were lost. Each level was treated separately and independently. The interactions were missing when the results were simply compared at
each level. Therefore, the HLM analyses are employed so that the analyses of different levels can be incorporated into one system. Further, the results of the two different approaches - descriptive and HLM model-based analyses will be compared to address the question of whether methodology matters when addressing the issue of achievement gap.

Other than investigating the trends in achievement and achievement gap at different levels, the analyses conducted in this section provided a base line for comparison to the HLM results. The comparison will allow us to look at the more complex student growth which is simultaneously linked with school/district characteristics in HLM models. It provides the justification for further investigations.

### Section Two: Hierarchical Linear Modeling

Hierarchical Linear Modeling (HLM) is often applied to study the overall structure of growth over time and to assess the extent to which individuals vary in their growth patterns. Growth in this study includes growth intercept (initial status) and growth rate (rate of change). The terminology of growth here does not necessarily indicate that achievement scores increased from academic year 2002-03 to 2004-05. Hence, growth intercept and initial status as well as growth rate and rate of change are exchangeable in this study.

In addition to the overall structure of growth, ethnicity is included in the HLM models to investigate ethnic differences in achievement growth, which is also called the achievement growth gap. Moreover, to address the hierarchical data structure that students are nested within schools and school districts, the growth patterns of schools and
school districts are further examined. Covariates are included to investigate what individual and organizational characteristics are associated with growth intercept, growth rate, and achievement growth gap.

Four models are presented to answer different research questions. They are 1) two-level unconditional model, 2) two-level conditional model, 3) three-level unconditional model, and 4) three-level conditional model. Unconditional models, without any covariates, are first presented to give an overview of model structure. Independent variables are then added to investigate the ethnic differences in achievement growth when holding other predictors constant.

This section is organized into three parts. First, preliminary analyses, including the two-level unconditional model and the three-level unconditional model; Second, ethnic differences in achievement growth between Hispanic and White students, including two-level and three-level models; Third, ethnic differences in achievement growth between Black and White students, including two-level and three-level models.

**Preliminary Analyses**

**Two-Level Unconditional Model**

The two-level unconditional model is introduced when no independent variables (or covariates) are included except the covariate of time. In this model, math achievement is only the function of time (academic year). Thus, we can investigate whether the achievement growth parameters vary across individuals. The model is expanded as follows:
Level 1 is an individual growth model of academic achievement at time $t$ of student $i$:

$$Y_{ni} = \pi_{0i} + \pi_{1i} (ACADEMIC.YEAR) + e_{ni}, \quad e_{ni} \sim N(0, \sigma^2)$$

where

$(ACADEMIC.YEAR)_i$ is 0 for the 2002-03 academic year, 1 for 2003-04, 2 for 2004-05;

$\pi_{0i}$ is the initial math score of student $i$ at the starting academic year;

$\pi_{1i}$ is the growth rate for student $i$ during the three academic years.

Level 2, the person-level model investigated whether there is statistically meaningful random variation in $\pi_0$ and $\pi_1$.

$$\pi_{0i} = \beta_{00} + r_{0i},$$
$$\pi_{1i} = \beta_{10} + r_{1i},$$

where

$\beta_{00}$ is the mean initial status (intercept)

$\beta_{10}$ is the mean growth rate over three years

$r_{0i}$ and $r_{1i}$ are assumed to be multivariate normally distributed, both with expected values of 0. We label these variances as

$$Var(r_{0i}) = \tau_{00},$$
$$Var(r_{1i}) = \tau_{11},$$

and the covariance between them as

$$Cov(r_{0i}, r_{1i}) = \tau_{01}$$
Table 4.2.1
Two-Level Unconditional Linear Model of Growth in Math Achievement

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>se</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean initial status, $\beta_{00}$</td>
<td>528.4</td>
<td>0.6</td>
<td>.000</td>
</tr>
<tr>
<td>Mean growth rate, $\beta_{10}$</td>
<td>20.8</td>
<td>0.2</td>
<td>.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Variance Component</th>
<th>Df</th>
<th>$x^2$</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Status, $r_{0i}$</td>
<td>5143.4</td>
<td>14904</td>
<td>202201.2</td>
<td>.000</td>
</tr>
<tr>
<td>Growth rate, $r_{1i}$</td>
<td>92.9</td>
<td>14904</td>
<td>20844.8</td>
<td>.000</td>
</tr>
<tr>
<td>Level-1 variation, $e_{ii}$</td>
<td>459.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reliability of OLS Regression Coefficient Estimate

<table>
<thead>
<tr>
<th>Initial status, $\pi_{0i}$</th>
<th>0.918</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth rate, $\pi_{1i}$</td>
<td>0.270</td>
</tr>
</tbody>
</table>

**Fixed Effects**

The estimated mean initial status ($\beta_{00}$) and the estimated mean growth rate ($\beta_{10}$) is 528.4 and 20.7, meaning that the mean score in the starting academic year (2002-03) for all the students was 528.4 and that the mean score, on average, increased by 20.8 points per year. The standard errors for the mean growth parameters (0.6 & 0.2) are small. The estimations of true mean initial status and growth rate fall into a narrow range due to the big sample size.

**Random Effects**

Random effect here refers to the individual variation in growth trajectories including growth intercept $\pi_{0i}$ and growth rate $\pi_{1i}$. The estimates of the variance of individual growth intercept and growth rate are 5143.4 and 92.9, respectively. The $x^2$ tests for both parameters are significant to reject the null hypothesis that the individuals
do not vary with regard to their growth intercept and growth rate. Therefore, we can conclude that there is significant variation in the math achievement scores at the starting academic year as well as in how fast the scores increased. There is a ratio of 50 or so between the estimate of the variance of individual growth intercept (5143.4) and that of growth rate (92.9), indicating that there is much larger variability in the initial status than in the growth rate.

For the growth intercept, scores scatter around the mean initial status 528.4 with standard deviation \((5143.4)^{1/2} \approx 71.7\). Hence, 95% of the starting year scores scatter between 385.0 and 671.8. For the growth rate, scores scatter around the mean growth rate 20.8 with standard deviation \((92.9)^{1/2} \approx 9.6\). Hence, 95% of the growth rates fall between 1.6 and 40.0. The large ranges in both growth intercept and growth rate show that students differ a great deal from each other in terms of their initial scores in 2002-03 and how fast the initial scores increased. For example, a child whose growth intercept is one standard deviation above average is expected to have the starting year score of 528.4+71.7=600.1 points, and a child whose growth rate is one standard deviation above average is expected to increase 20.8+9.6=30.4 points per year.

Reiability

The estimates of reliability of growth intercept and growth rate are 0.918 and 0.270, respectively. The results indicate there is substantial individual difference (parameter variance) in terms of the growth intercept. However, low reliability of growth rate suggests a considerable error variance in the estimation. One possible explanation is that compared with growth intercept (initial scores at the starting academic year), there is
less parameter variance in growth rate. Stated differently, compared with their initial scores, students differed less in their growth rate.

**Correlations of Growth Rate with Growth Intercept**

One advantage of the HLM analyses is that consistent estimates of the correlation of growth intercept and growth rate can be obtained.

\[
\hat{\rho} = \frac{-476.8}{(5143.4*92.9)^{1/2}} = - 0.69
\]

The estimated correlation coefficient \( \hat{\rho} (-0.69) \) indicates that there is strong negative correlation between growth intercept and growth rate. Therefore, the greater the starting-year achievement scores, the smaller the rate of growth over the three years. The relationship between the two parameters implies the necessity of controlling for the intercept when the growth rates of different ethnic groups are compared.

**Three-Level Unconditional Model (Level Three - School)**

When students are nested in schools and districts, a primary research question is how individual growth trajectories vary with organizational context. The three-level unconditional HLM model is applied to provide a perspective on whether the ethnic differences in achievement growth vary substantially across schools or districts. The three levels include:

1. **level 1** - an individual growth model of the academic achievement at time \( t \) of student \( i \) in school \( j \):

\[
Y_{ij} = \pi_{0ij} + \pi_{ij} (ACADEMIC.YEAR)_{ij} + e_{ij}, \quad e_{ij} \sim N (0, \sigma^2)
\]

2. **level 2** - a person-level model,
\[ \pi_{0ij} = \beta_{00j} + r_{0ij}, \]
\[ \pi_{1ij} = \beta_{10j} + r_{1ij}, \]

and, (3) level 3 - a school-level or district-level model (schools or districts are the third level units in separate models),

\[ \beta_{00j} = \gamma_{000} + u_{00j}, \]
\[ \beta_{10j} = \gamma_{100} + u_{10j}, \]

Where

- \( \beta_{00j} \) represents the mean initial status within school/district \( j \)
- \( \gamma_{000} \) is the average mean initial status across schools/districts
- \( \beta_{10j} \) is the mean growth rate of three academic years within school/district \( j \)
- \( \gamma_{100} \) is the average mean growth rate of three academic years across schools/districts

(Note that the notations are fully explained in the chapter on research design.)

Table 4.2.2 displays the fixed and random effects of the three-level unconditional model when the third level units are schools.

The above table showed that the average math achievement scores of all schools increased. The average scores of all schools (\( \gamma_{000} \)) started at 526.4 in year 2002-03 and increased at the rate of 21.6 points per year (\( \gamma_{100} \)). These two coefficients were close to the individual mean initial score (528.4) and individual mean growth rate (20.8) (see Table 4.2.1), suggesting that students are roughly evenly distributed across schools as to how high they scored in year 2002-03 as well as how fast their scores increased across the three years. The small standard errors for the two growth parameters (3.3 & 0.6) indicated that the true initial status and growth rate fell into a relatively narrow range.
Table 4.2.2
*Three-Level Unconditional Model of Growth in Math Achievement across Schools*

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>se</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average initial status across schools, $γ_{000}$</td>
<td>526.4</td>
<td>3.3</td>
<td>.000</td>
</tr>
<tr>
<td>Average growth rate across schools, $γ_{100}$</td>
<td>21.6</td>
<td>0.6</td>
<td>.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance</th>
<th>df</th>
<th>$\chi^2$</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 (scores)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variation, $e_{ij}$</td>
<td>447.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2 (between students within schools)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual initial status, $r_{0ij}$</td>
<td>4256.6</td>
<td>14807</td>
<td>57459.6</td>
<td>.000</td>
</tr>
<tr>
<td>Individual growth rate, $r_{1ij}$</td>
<td>62.0</td>
<td>14807</td>
<td>18318.9</td>
<td>.000</td>
</tr>
<tr>
<td>Level 3 (between schools)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School average initial status, $μ_{00j}$</td>
<td>1011.2</td>
<td>97</td>
<td>2925.3</td>
<td>.000</td>
</tr>
<tr>
<td>School average growth rate, $μ_{10j}$</td>
<td>35.2</td>
<td>97</td>
<td>1607.7</td>
<td>.000</td>
</tr>
</tbody>
</table>

The second panel of Table 4.2.2 showed that the variance in initial status and growth rate was decomposed into within- and between-school components. Significant variation was found within schools (among students) for individual initial status and individual growth rates ($r_{0ij}$ and $r_{1ij}$) as well as between schools for school average initial status and school average growth rates ($μ_{00j}$ and $μ_{10j}$). By comparing the $\chi^2$ statistics accompanying these variance components, one can see that the variations in initial status and growth rates between schools were both much smaller than the variations within schools.
Based on the variance component estimates, we can compute the percentage of variation that lies between schools in both initial status and growth rate. Specifically, percent of variance between schools in initial status is

\[
\frac{\tau_{\beta 00}}{\tau_{\beta 00} + \tau_{\pi 00}} = \frac{1011.2}{1011.2 + 4256.6} = 0.19,
\]

and percent of variance between schools in growth rate is

\[
\frac{\tau_{\beta 11}}{\tau_{\beta 11} + \tau_{\pi 11}} = \frac{35.2}{35.2 + 62.0} = 0.36.
\]

About 19% of the variance in initial status (achievement scores in academic year 2002-03) lies between schools. This is consistent with school effects in previous studies where 10% to 30% of the achievement variability was found between schools. The result for growth rates, however, is surprising: Almost 36% of the variance is between schools. This indicated that not only did schools in the study differ in their initial average achievement scores in year 2002-03, they differed even more in their average growth rates. In order to investigate school effects, some school characteristics were included later in this section to explain the variability in student achievement growth.

Another approach to examining the within-school and between-school effects is to decompose the correlation between initial status and growth rate into within-(level 2) and between-school (level 3) components. The results below show that within a typical school, the estimated correlation between the two growth parameters (initial status and growth rate) is -0.69. It means that student initial achievement scores in 2002-03 are strongly associated with their growth rates across the three years. The correlation is
slightly weaker at the school level (-0.66). The average initial scores of all schools are still strongly correlated with the average growth rates of these schools.

Variance-covariance components and correlations at Level-2 and Level-3 are listed below:

\[
\begin{bmatrix}
4256.6 & -0.69 \\
-356.4 & 62.0
\end{bmatrix} = \tilde{T}_x = \begin{bmatrix}
\tilde{\tau}_{x11} \\
\tilde{r}_{x12} \\
\tilde{\tau}_{x22}
\end{bmatrix}
\]

\[
\begin{bmatrix}
1011.2 & -0.66 \\
-125.0 & 35.2
\end{bmatrix} = \tilde{T}_z = \begin{bmatrix}
\tilde{\tau}_{\beta11} \\
\tilde{r}_{\beta12} \\
\tilde{\tau}_{\beta22}
\end{bmatrix}
\]

In sum, the three-level unconditional model provided important statistics for studying individual growth, including the apportioning of variability in the individual growth parameters at different levels, the variance components, and the correlations between growth parameters. In this particular application, the variance component decomposition highlighted an important feature of the data: the high percentage of variation in growth rates that lies between schools.

**Three-Level Unconditional Model (Level Three - District)**

When districts, instead of schools, serve as the third level units in the HLM model, the results are slightly different. Table 4.2.3 compared the three-level unconditional model with schools as the third level units with the model with districts as the third level units. It showed that the average initial score in year 2002-03 of all districts (521.7) was slightly smaller than the one of all schools (526.4), indicating that students were roughly evenly distributed across schools and districts as to their initial scores. If the average score of districts is significantly lower than the score of schools, it implies that low performing students tend to cluster in relatively small districts.
The average growth rate of all districts (22.5) is slightly smaller than the one of all schools (21.6). This indicates that students were roughly evenly distributed across schools and districts as to their growth rates. If the average growth rate of districts is significantly higher than that of schools, it suggests that students with higher growth rates tend to cluster in relatively small districts.

The percent of variance in initial status between districts is much lower (9%) than the one between schools (19%), indicating that districts did not differ as much as schools with regard to their average scores in year 2002-03. Similar to the model for schools, a larger percent of variance (20%) in growth rates were between districts than the one in initial status (9%), suggesting that districts might differ from each other more in growth rate than in their average achievement in year 2002-03. It is also noticeable that compared with districts (20%), a larger percent of variance in growth rate lies between schools (36%). Schools may account more for the variation in student growth rate.

Table 4.2.3
The Comparison between the Three Level Unconditional Models with Schools and Districts as the Third Level Units, Respectively

<table>
<thead>
<tr>
<th>Parameters in 3-level HLM models</th>
<th>Schools</th>
<th>Districts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average initial status</td>
<td>526.4</td>
<td>521.7</td>
</tr>
<tr>
<td>Average growth rate</td>
<td>21.6</td>
<td>22.5</td>
</tr>
<tr>
<td>% of variance in initial status between organizations</td>
<td>19%</td>
<td>9%</td>
</tr>
<tr>
<td>% of variance in growth rate between organizations</td>
<td>36%</td>
<td>20%</td>
</tr>
</tbody>
</table>
The Hispanic-White Gap in Achievement Growth

Two-Level Conditional Model

The two-level unconditional model above focuses on the growth parameters when only the time covariate is included in the model. By adding other variables, especially the covariates of ethnicity, the two-level conditional models are applied to investigate: (1) whether the achievement scores of minority students (Hispanics and Blacks) increased faster on average than the scores of White students, (2) whether growth rates were significantly correlated with growth intercepts (initial statuses), and (3) to what extent the differences in growth parameters can be accounted for by student characteristics. To address different questions, different models are employed.

Model 1 (Two-Level Conditional Model with the Effect of Ethnicity)

The dummy variable of ethnicity (Hispanic/Black=0 White=1) is used to investigate achievement growth gap, which refers to ethnic differences in achievement growth intercept (initial status) and growth rate (rate of change).

The level-1 model remains the same as the one in the unconditional model. The level-2 model represents the variability in growth intercept and growth rate. For the achievement growth differences between Hispanic and White students, the level-2 model is formulated as follows:

\[ \pi_{0i} = \beta_{00} + \beta_{01}(ETHNICITY_{Hi-W})_i + r_{0i} \]

\[ \pi_{1i} = \beta_{10} + \beta_{11}(ETHNICITY_{Hi-W})_i + r_{1i} \]

Compared with the unconditional model, the additional parameters in this model are \( \beta_{01} \) and \( \beta_{11} \). \( \beta_{01} \) represents the difference in growth intercept (initial status) between
Hispanic and White students. $\beta_{11}$ represents the difference in growth rates between Hispanic and White students.

The table below demonstrates the estimated the fixed effect and the random effect of the two-level conditional model with the effect of ethnicity.

Table 4.2.4
Two-Level Conditional Model of Growth in Math Achievement with the Effect of Ethnicity

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>se</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for initial status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Hispanic score, $\beta_{00}$</td>
<td>485.2</td>
<td>1.3</td>
<td>.000</td>
</tr>
<tr>
<td>Mean ethnicity contrast, $\beta_{01}$</td>
<td>58.4</td>
<td>1.5</td>
<td>.000</td>
</tr>
<tr>
<td>Model for growth rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Hispanic rate, $\beta_{10}$</td>
<td>26.2</td>
<td>0.4</td>
<td>.000</td>
</tr>
<tr>
<td>Mean ethnicity contrast, $\beta_{11}$</td>
<td>-7.0</td>
<td>0.4</td>
<td>.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Variance Component</th>
<th>df</th>
<th>$x^2$</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Status, $r_{0i}$</td>
<td>4452.6</td>
<td>12879</td>
<td>164627.6</td>
<td>.000</td>
</tr>
<tr>
<td>Growth rate, $r_{1i}$</td>
<td>83.6</td>
<td>12879</td>
<td>17720.5</td>
<td>.000</td>
</tr>
<tr>
<td>Initial status * Growth rate $r_{01}$</td>
<td>-406.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level-1 variation, $e_{ii}$</td>
<td>451.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The coefficients of ethnicity ($\beta_{01}$ & $\beta_{11}$) were found to be significantly related to math achievement (p<.01). On average, White students started 58.4 points higher than Hispanic students ($\beta_{01} = 58.4$). Thus, since the average math score of Hispanic students in year 2002-03 was 485.2, the average starting-year score of White students was 543.6 ($= 485.2 + 58.4$) points.
As to the individual growth rate, the scores of White students, on average, increased at a slower rate compared with their Hispanic peers ($\beta_{11} = -7.0$). Thus, when the scores of Hispanic students increased, on average, at a rate of 26.2 points per year, the scores of White students increased, on average, at a rate of 19.2 ($= 26.2 - 7.0$) points per year.

The ethnic differences in both growth intercept and growth rate were significant. The results showed that achievement growth gap existed between Hispanic and White students in terms of their math achievement.

As to the random effects, the estimates for the variance of individual growth intercept $r_{0i}$ and growth rate $r_{1i}$ were 4452.6 and 83.6, respectively. They both were found significant, indicating that individuals still varied significantly with regard to their initial scores and rates of change after their ethnicity were held constant.

As to the correlation between initial status and growth rate, $\hat{\rho}$ (-0.67) indicated that there was strong negative correlation between growth intercept and growth rate (see the calculation below). Thus, the greater the starting-year achievement scores, the smaller the rate of growth over the three years. The strong relationship between the two growth parameters validates the necessity of controlling for growth intercept when growth rates are compared and discussed.

\[
\hat{\rho}(\tau_{0i}, \tau_{1i}) = \frac{\hat{\epsilon}_{01}}{(\hat{\tau}_{00}\hat{\tau}_{11})^{1/2}} = \frac{-406.4}{(4452.6 * 83.6)^{1/2}} = -0.67
\]

Compared with two-level unconditional model, the strength of the correlation between the two growth parameters in this model is very close to the one in the two-level
unconditional model (-0.69). It implies that after taking into account student ethnic background, the relationship between where they started and how fast their scores increased did not change much.

In order to understand how much more variance can be accounted for by exploring a new model, the concept of proportion reduction in variance is applied by comparing the level 2 variance $\tau_{00}$ and $\tau_{11}$ between the two models.

Proportion of variance explained in initial status $\beta_{0i}$

$$= \frac{\hat{\tau}_{00} (\text{Model Unconditional}) - \hat{\tau}_{00} (\text{Model 1})}{\hat{\tau}_{00} (\text{Model Unconditional})} = \frac{5016.5 - 4452.6}{5016.5} = 0.11$$

Proportion of variance explained in growth rate $\beta_{1i}$

$$= \frac{\hat{\tau}_{11} (\text{Model Unconditional}) - \hat{\tau}_{11} (\text{Model 1})}{\hat{\tau}_{11} (\text{Model Unconditional})} = \frac{92.5 - 83.6}{92.5} = 0.10$$

The table below displays how much variance can be explained in initial status and growth rate by adding ethnicity to the unconditional model. Ethnicity can account for about 11% of variance in initial status and about 10% of variance in growth rate in the student math achievement scores.

<table>
<thead>
<tr>
<th>Model</th>
<th>Initial Status Var ($\pi_{0i}$)</th>
<th>Growth Rate Var ($\pi_{1i}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconditional</td>
<td>5016.5</td>
<td>92.5</td>
</tr>
<tr>
<td>Model 1 (Conditional on ethnicity)</td>
<td>4452.6</td>
<td>83.6</td>
</tr>
<tr>
<td>Proportion of variance explained</td>
<td>0.11</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Model 2 (Holding Initial Status Constant)

The strong negative correlation between growth intercept and growth rate provides an excellent rationale for including the growth intercept in the model for comparing the growth rates of different ethnic groups. It is likely that minority students have lower starting points and thus tend to grow faster (i.e. regression toward the mean) than their non-minority counterparts. On the other hand, the tests may be constructed in a way that within some score ranges, items may be too difficult or too easy to differentiate test-takers effectively. Thus, the location of the scores may be related with how fast the scores can grow. When the change may not truly reflect the improvement of academic performance, it is misleading to simply compare growth rates. Regressions toward the mean, as well as the location of the scores, need to be controlled for so that we can make a valid conclusion when the growth rates of different ethnic groups are compared. The solution in this study is to hold the initial status constant.

Therefore, when the level-one model remains the same, the level-two model here is illustrated as

\[ \pi_{0i} = \beta_{00} + \beta_{01}(ETHNICITY_{H-W})_i + r_{0i} \]

\[ \pi_{1i} = \beta_{10} + \beta_{11}(ETHNICITY_{H-W})_i + \beta_{12}(Initial\,status - \overline{Initial\,status}) + r_{1i} \]

When the model for initial status \( \pi_{0i} \) remains the same, the initial status is added into model for growth rate. The initial status is grand mean centered so that other effects are easier to interpret. In the grand mean centered model, the explanatory variables are centered around the overall mean. In this model between the choices of grand mean centering or non-centering for initial status, only the slope intercept \( \beta_{10} \) was found to be
different while other fixed effects and random effects were stable. Without centering, the slope intercept $\beta_{10}$ was 68.9 when with grand centering it was 22.1. Since the coefficient for initial score ($\beta_{12}$) is -0.0878 (see Table 4.2.6), the grand mean can be calculated by using the difference between the two intercept values.\footnote{(68.9-22.1)/0.0897=533.0}

Table 4.2.6

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>se</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for initial status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Hispanic score, $\beta_{00}$</td>
<td>485.2</td>
<td>1.3</td>
<td>.000</td>
</tr>
<tr>
<td>Mean ethnicity contrast, $\beta_{01}$</td>
<td>58.7</td>
<td>1.5</td>
<td>.000</td>
</tr>
<tr>
<td>Model for growth rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Hispanic rate, $\beta_{10}$</td>
<td>22.1</td>
<td>0.3</td>
<td>.000</td>
</tr>
<tr>
<td>Mean ethnicity contrast, $\beta_{11}$</td>
<td>-1.9</td>
<td>0.4</td>
<td>.000</td>
</tr>
<tr>
<td>Initial score, $\beta_{12}$</td>
<td>-0.1\footnote{-0.1 here is the rounding result of - 0.0878}</td>
<td>0.003</td>
<td>.000</td>
</tr>
</tbody>
</table>

Another notable feature about this model was that there was no significant variance for the slope of individual growth rate ($r_{1i}$). The addition of the predictor of initial status led to a sudden large increase in the number of iterations required for convergence. When too many iterations are required to ensure that convergence has been reached, the slope associated with the problem element of tau needs to be fixed in order to obtain a stable solution. In other words, a common and fixed slope over units is adequate. Therefore, the random coefficient for the slope of growth rate is removed, implying that growth rate does not vary significantly from individual to individual while initial status is controlled for. The estimate for the variances of individual growth

\footnote{(68.9-22.1)/0.0897=533.0}
\footnote{-0.1 here is the rounding result of - 0.0878}

126
intercept $r_{0i}$ is 4486.4, which is very close to the variance in the previous model (model 1). Holding initial score constant for growth rate has no impact on the variation of initial scores.

Table 4.2.6 presents the results of fixed effects of conditional model with initial status controlled for. The table showed that the covariates of ethnicity ($\beta_{01}$ & $\beta_{11}$) were still strongly related to math achievement ($p<.01$), after initial scores were controlled for. On average, White students started 58.7 points higher than Hispanic students ($\beta_{01}$), which was close to the value in model 1 (58.4). As to the individual growth rate, the test scores of White students increased, on average, at a rate of 1.9 points per year slower than their Hispanic peers ($\beta_{11}$). The difference was much smaller than the one in model 1 (7.0 points) when initial status was not controlled for. Therefore, when initial scores were controlled for, although Hispanic students still grew faster in terms of their math achievement scores than their White peers, the growth gap was not statistically significant any more. If White students and Hispanic students started with the same scores, Hispanic students only grew faster by approximately 2 points per year than their White counterparts. To decrease the achievement gap, the growth gap was expected to be large so that minority students can grow faster than their White peers over time. The results suggest that the difference in growth rates is small compared to the average difference in starting values, so that the achievement gap will narrow very slowly.

The covariate of initial status $\beta_{12}$ (-0.1) indicates that after student ethnicity being held constant, when initial score increases one point, growth rate would decrease 0.1
point. For example, if two students are of the same ethnicity, when student A’s initial score is 10 points higher than student B, student A is expected to grow 1 point per year slower than student B. However, the model cannot tell whether the correlation between initial status and growth rate is constant across ethnic groups. Stated differently, it is of concern whether Hispanic students and White students differ from each other as to the relationship between the two growth parameters. Model 3 is applied then when the interaction effect between ethnicity and initial status being held constant at the level-two model for growth rate.

**Model 3 (Holding the Interaction Effect between Ethnicity and Initial Status Constant)**

The previous models have found that on average Hispanic students scored lower than their White peers in terms of their starting year scores (initial status). Hence, when the growth rates were found to differ significantly between the two ethnic groups, the gap can be due to the difference in initial status instead of in ethnicity. In order to control for the interaction between ethnicity and initial status, the interaction effect is included into the second level model for growth rate. The model for intercept remains the same.

Therefore, for Model 3, the second level is as follows:

\[
\pi_{0i} = \beta_{00} + \beta_{01}(ETHNICITY_{H-W})_i + r_{0i}
\]

\[
\pi_{1i} = \beta_{10} + \beta_{11}(ETHNICITY_{H-W})_i + \beta_{12}(Initial.status - \overline{Initial.status}) + \\
\beta_{13}(Ethnicity * Initial - Ethnicity * Initial)
\]

where \( \beta_{13} \) is the interaction effect (grand centered) between ethnicity and initial status when other covariates are held constant.
Table 4.2.7
*Fixed Effects of the Two-Level Model after Initial Status and the Interaction Effect between Ethnicity and Initial Status Being Controlled for (Compared with Model 2)*

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Model 2 Coefficient</th>
<th>Model 2 p value</th>
<th>Model 3 Coefficient</th>
<th>Model 3 p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for initial status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Hispanic Score, $\beta_{00}$</td>
<td>485.2</td>
<td>.000</td>
<td>485.2</td>
<td>.000</td>
</tr>
<tr>
<td>Mean ethnicity contrast, $\beta_{01}$</td>
<td>58.7</td>
<td>.000</td>
<td>58.7</td>
<td>.000</td>
</tr>
<tr>
<td>Model for growth rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Hispanic rate, $\beta_{10}$</td>
<td>22.1</td>
<td>.000</td>
<td>21.5</td>
<td>.000</td>
</tr>
<tr>
<td>Mean ethnicity contrast, $\beta_{11}$</td>
<td>-1.9</td>
<td>.000</td>
<td>-1.0</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Initial score, $\beta_{12}$</td>
<td>-0.1</td>
<td>.000</td>
<td>-0.1</td>
<td>.000</td>
</tr>
<tr>
<td>Ethnicity x Initial scores $\beta_{13}$</td>
<td>-0.002</td>
<td>&gt;.05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Compared with model 2, the prediction of initial status on growth rate remains significant, suggesting that initial status is a more important predictor than ethnicity. The result showed that whether students were Hispanic or White did not matter as much as where their initial scores were.

**Model 4 (Holding Other Student-Level Variables Constant)**

Other student variables such as gender (Male=0, Female=1) and student mobility$^{10}$ were also investigated to see whether they were significantly correlated with initial status and growth rate. To be more specific, it is of interest to investigate the differences in achievement growth between male and female students as well as among students who did not move and who moved for different reasons. Meanwhile, the two predictors, as well as their interactions with ethnicity, served as covariates so that we can detect the ethnic differences in achievement growth when gender and mobility were held constant.

---

$^{10}$ Non-mover=0, Bureaucratic Mover=1, Mover (Once)=2, Mover (Twice)-A=3, Mover (Twice)-B=4
Therefore, for Model 4, the second level model is illustrated as follows:

\[ \pi_{0i} = \beta_{00} + \beta_{01}(ETHNICITY_{H-W}), + \beta_{02}(Gender) + \beta_{03}(Mobility) + \beta_{04}(Ethnicity * Gender) + \beta_{05}(Ethnicity * Mobility) + \epsilon_{0i} \]

\[ \pi_{1i} = \beta_{10} + \beta_{11}(ETHNICITY_{H-W}), + \beta_{12}(Gender) + \beta_{13}(Mobility) + \beta_{14}(Initial \text{ status} - \text{Initial \text{ status}}) + \beta_{15}(Ethnicity * Initial - Ethnicity * Initial) + \beta_{16}(Ethnicity * Gender) + \beta_{17}(Ethnicity * Mobility) \]

Table 4.2.8
Fixed Effects of the Two-Level Conditional Model with Person-Level Predictors (Compared with Model 3)

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Model 3 Coefficient</th>
<th>p value</th>
<th>Model 4 Coefficient</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for initial status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Hispanic score, ( \beta_{00} )</td>
<td>485.2</td>
<td>.000</td>
<td>486.8</td>
<td>.000</td>
</tr>
<tr>
<td>Mean ethnicity contrast, ( \beta_{01} )</td>
<td>58.7</td>
<td>.000</td>
<td>64.8</td>
<td>.000</td>
</tr>
<tr>
<td>Gender (Male=0), ( \beta_{02} )</td>
<td></td>
<td></td>
<td>4.2</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Mobility, ( \beta_{03} )</td>
<td>-6.3</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity x Gender, ( \beta_{04} )</td>
<td>-7.3</td>
<td>.015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity x Mobility, ( \beta_{05} )</td>
<td>-9.8</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model for growth rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Hispanic rate, ( \beta_{10} )</td>
<td>21.5</td>
<td>.000</td>
<td>23.0</td>
<td>.000</td>
</tr>
<tr>
<td>Mean ethnicity contrast, ( \beta_{11} )</td>
<td>-1.0</td>
<td>&gt;.05</td>
<td>-2.9</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Gender (Male=0), ( \beta_{12} )</td>
<td>-1.8</td>
<td>.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobility (Non-mover), ( \beta_{13} )</td>
<td>-1.0</td>
<td>.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial score, ( \beta_{14} )</td>
<td>-.1</td>
<td>.000</td>
<td>-.09</td>
<td>.000</td>
</tr>
<tr>
<td>Ethnicity x Initial scores, ( \beta_{15} )</td>
<td>-0.002</td>
<td>&gt;.05</td>
<td>-0.002</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Ethnicity x Gender, ( \beta_{16} )</td>
<td>2.3</td>
<td>.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity x Mobility, ( \beta_{17} )</td>
<td>1.2</td>
<td>.002</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Based on table 4.2.8, White students, on average, scored 64.8 points (\( \beta_{01} \)) higher than their Hispanic peers in year 2002-03 when student gender, mobility, and their interactions with ethnicity were held constant. Compared with model 3, the addition of
person-level predictors widened the ethnic gap in initial scores (58.7 points in model 3). The result implies the achievement gap in year 2002-03 could be even larger if the other student characteristics were held constant.

The main effects of gender and mobility need to be interpreted with their interaction effects with ethnicity since they both were found statistically significant. As to the gender difference, without including the interaction between gender and ethnicity, female students, on average, were found to score slightly lower than their male peers in their initial scores but had slightly higher growth rate. These differences were not statistically significant. When the interaction between ethnicity and gender was included, the significant interaction suggested that the gender difference was not constant across ethnicity. We found that irrespective of gender, Hispanic students scored much lower initially but grew faster than their White peers. The gender differences were not statistically significant except that Hispanic female students had a significantly higher average growth rate than Hispanic male students.

Mobility was confounded with ethnicity too. In general, as to the initial scores, students who never moved scored higher than those who moved; students who moved once scored higher than those who moved twice; and students who moved due to bureaucratic reason scored higher than those who moved for other reasons. No significant difference in growth rate was found among the different types of movers until the interaction between mobility and ethnicity was added in the model. Students with less mobility had higher average growth rates. No matter of student mobility, Hispanic students were found to score lower in initial scores but grew faster than their White peers.
As to the model for growth rate, ethnicity was found not significantly related with growth rate ($\beta_{11}$). There was no significant ethnic difference between Hispanic students and White students in how fast their math scores increased across the three years, when initial status, the interaction of initial status with ethnicity, gender, and student mobility were held constant. In sum, how fast the scores increased, to a large extent, depended on where students started. No significant ethnic difference in growth rate was found after growth intercept was taken into account.

Summary

To investigate whether Hispanic students and White students had different growth patterns when their initial scores were held constant, the expected growth rates for both groups of students were calculated when their initial scores were 1) 100 points below the overall mean score, 2) the overall mean score, and 3) 100 points above the overall mean score. The expected growth rates are presented in Table 4.2.9.

Table 4.2.9

<table>
<thead>
<tr>
<th>Initial Status</th>
<th>Hispanic (points/year)</th>
<th>White (points/year)</th>
<th>Difference (growth rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 points below the student mean score</td>
<td>35.8</td>
<td>29.2</td>
<td>6.6</td>
</tr>
<tr>
<td>Student mean score</td>
<td>25.8</td>
<td>19.0</td>
<td>6.8</td>
</tr>
<tr>
<td>100 points above the student mean score</td>
<td>15.8</td>
<td>8.8</td>
<td>7.0</td>
</tr>
</tbody>
</table>

Based on the two-level conditional model with initial status controlled for (see Table 4.2.6), the average growth rates for students starting with mean score (528.4) is 25.8 ($=21.5-(528.4-485.2)\times(-0.1)$) for Hispanic students and 19.0 ($=21.5-1.0-[(58.7-(528.4-485.2))\times(-0.1)]$) for White students. For the students who started 100 points below
the mean score, if they are Hispanic, their growth rate is 35.8 (=25.8-(-0.1x100)); if they are white, their growth rate is 29.2 (=19.0-(-0.1x100)-(-0.002x100)). For students who started 100 points above the mean score, if they are Hispanic, their growth rate is 15.8 (=25.8+ (-0.1x100)); if they are white, their growth rate is 8.8 (=19.0+(-0.1x100)+(-0.002x100)).

The figure below illustrates the expected growth trajectory for Hispanic and White students based on their initial status. It shows that although White students on average grew slower than Hispanic students, initial score was a even more important predictor of growth rate. Starting from different scores, to a large extent, decides how fast students grow and what their later achievement scores are.

![Figure 4.2.1](image-url)  
*Figure 4.2.1. The Expected Growth Trajectory for Hispanic and White Students Based on Their Initial Status.*

Therefore, the approach of Hierarchical Linear Modeling not only provides us with a growth perspective on student achievement by displaying growth rates of different
ethnic groups but it also helps to disentangle the relationship between where students start and how fast they grow. Without considering this relationship, we may simply conclude that Hispanic students grew faster than White students in terms of their math achievement. However, when initial status was included to model individual growth rate, the ethnic gap in growth rate minimized.

**Three-Level Conditional Model (Level Three - School)**

Finding that there was a significant amount of variation in initial status and growth rate across schools and districts, we now include some explanatory variables to answer the research question - which school or district factors are associated with such variation? The level 1 model remains the same as the one in the three level unconditional models. At level 2, the dummy variable ethnicity (Hispanic = 0, White = 1) is included to examine the hypothesis that ethnicity is significantly related to the variability in initial status and growth rate. The achievement growth between the two ethnic groups can then be compared. Other individual-level variables, such as gender and mobility, are not included since the two variables were found not significantly related with growth rate in the two-level conditional model.

At level three, the variability among schools/districts in the growth parameters is addressed. The explanatory school variables include the minority (Hispanic) percentage in a school, school mobility rate, total student number, pupil-teacher ratio, and percentage of students who had Free or Reduced-price Lunch (FRL). The table below displays the fixed effects of student ethnicity and school variables on math achievement.
<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>s.e.</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>For Intercept</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For Intercept (initial status)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, $r_{000}$</td>
<td>496.7</td>
<td>2.5</td>
<td>.000</td>
</tr>
<tr>
<td>Minority percentage (Hispanic), $r_{001}$</td>
<td>-44.4</td>
<td>16.2</td>
<td>.008</td>
</tr>
<tr>
<td>Mobility rate, $r_{002}$</td>
<td>-100.5</td>
<td>20.3</td>
<td>.000</td>
</tr>
<tr>
<td>Total student number, $r_{003}$</td>
<td>0.02</td>
<td>.006</td>
<td>.004</td>
</tr>
<tr>
<td>Pupil teacher ratio, $r_{004}$</td>
<td>0.2</td>
<td>1.3</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>FRL students (%), $r_{005}$</td>
<td>-28.5</td>
<td>16.8</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>For Ethnicity (achievement gap)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, $r_{010}$</td>
<td>43.6</td>
<td>2.9</td>
<td>.000</td>
</tr>
<tr>
<td>Minority percentage (Hispanic), $r_{011}$</td>
<td>12.3</td>
<td>17.3</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Mobility rate, $r_{012}$</td>
<td>-3.1</td>
<td>60.5</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Total student number, $r_{013}$</td>
<td>-.01</td>
<td>.008</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Pupil teacher ratio, $r_{014}$</td>
<td>-1.4</td>
<td>1.3</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>FRL students (%), $r_{015}$</td>
<td>-34.7</td>
<td>18.3</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>For Time Slope</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For Intercept (growth rate)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, $r_{100}$</td>
<td>25.6</td>
<td>0.7</td>
<td>.000</td>
</tr>
<tr>
<td>Minority percentage (Hispanic), $r_{101}$</td>
<td>13.0</td>
<td>5.2</td>
<td>.014</td>
</tr>
<tr>
<td>Mobility rate, $r_{102}$</td>
<td>6.5</td>
<td>7.3</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Total student number, $r_{103}$</td>
<td>-.006</td>
<td>.002</td>
<td>.002</td>
</tr>
<tr>
<td>Pupil teacher ratio, $r_{104}$</td>
<td>-.067</td>
<td>.3</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>FRL students (%), $r_{105}$</td>
<td>-5.5</td>
<td>4.4</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>For Ethnicity (achievement growth gap)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, $r_{110}$</td>
<td>-5.3</td>
<td>0.6</td>
<td>.000</td>
</tr>
<tr>
<td>Minority percentage (Hispanic), $r_{111}$</td>
<td>-6.9</td>
<td>3.0</td>
<td>.025</td>
</tr>
<tr>
<td>Mobility rate, $r_{112}$</td>
<td>7.1</td>
<td>8.9</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Total student number, $r_{113}$</td>
<td>0.002</td>
<td>.001</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Pupil teacher ratio, $r_{114}$</td>
<td>0.1</td>
<td>0.3</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>FRL students (%), $r_{115}$</td>
<td>10.3</td>
<td>3.7</td>
<td>.007</td>
</tr>
</tbody>
</table>
The last panel in Table 4.2.10 represents the concept of ethnic differences in achievement growth in this study, which is also called achievement growth gap. The minority percentage, while not significantly related with the achievement gap in year 2002-03, was found significantly related with the differences in growth rates between Hispanic and White students. The higher percentage of Hispanic students in a school, the smaller the growth gap the school tended to have. Combining with the result in panel 2 and 3, when the percentage of Hispanic students in a school was positively related with the school average growth rate, it was found that the difference in growth rate between Hispanic and White students was even smaller. One possible explanation is that ethnicity might be confounded with the other school factors such as student socio-economic background. Hence, the more Hispanic students in a school, the more alike students were with regard to their achievement in 2002-03 and their growth rates.

Percentage of FRL students was also significantly related with the growth gap between Hispanic and White students. The more low socio-economic students in a school, the larger the differences in growth rates between the two ethnic groups were. One possible explanation was that the low socio-economic students tended to have high growth rates, which led to the relatively large difference in growth rates. Other school characteristics, such as mobility rate, total student number, and pupil-teacher ratio were not found significantly related with the achievement growth gap.

In order to examine whether residual variance of student math achievement still remains to be explained, the variance is decomposed into three levels. Table 4.2.11 confirms that there was significant random variation at each level of the organizational
units (student at level two and school at level three). For example, schools in this study significantly varied from each other with regard to their average achievement in 2002-03, average achievement gap in 2002-03, average growth rate, and average achievement growth gap. When some argue that school has little effect on student achievement or growth, the random effects of schools showed in this study can be a strong argument that school does matter.

Table 4.2.11
Variance Decomposition from a Three-Level Analysis (Level 3 - School)

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Variance Component</th>
<th>df</th>
<th>( \chi^2 )</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporal variation, ( e_{ij} )</td>
<td>448.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2 (between students within schools)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual initial status, ( r_{ij} )</td>
<td>3792.6</td>
<td>12409</td>
<td>88908.0</td>
<td>.000</td>
</tr>
<tr>
<td>Individual growth rate, ( r_{ij} )</td>
<td>54.0</td>
<td>12409</td>
<td>15048.2</td>
<td>.000</td>
</tr>
<tr>
<td>Level 3 (between schools)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School average initial status, ( \mu_{00j} )</td>
<td>113.5</td>
<td>87</td>
<td>171.0</td>
<td>.000</td>
</tr>
<tr>
<td>School average growth rate, ( \mu_{10j} )</td>
<td>21.9</td>
<td>87</td>
<td>285.4</td>
<td>.000</td>
</tr>
<tr>
<td>School average initial gap, ( \mu_{01j} )</td>
<td>221.0</td>
<td>87</td>
<td>177.6</td>
<td>.000</td>
</tr>
<tr>
<td>School average growth gap, ( \mu_{11j} )</td>
<td>4.1</td>
<td>87</td>
<td>115.5</td>
<td>.022</td>
</tr>
</tbody>
</table>

Three-Level Conditional Model (Level Three - District)

When the third-level units were replaced by districts, the district characteristics on student achievement and achievement growth were then investigated. Table 4.13 displays the fixed effects of ethnicity and district variables on math achievement. For the initial status (see panel 1), pupil-teacher ratio was significantly related with the district average achievement score in year 2002-03, while it was not significantly related with the school
average achievement score (see Table 4.2.12, panel 1). One explanation can be that pupil teacher ratio varied more across districts than across schools. The percentage of English Language Learner (ELL) students was also found significantly related with the district average score in 2002-03. The higher the percentage of ELL students in a district, the district average score tended to be lower.

As to the achievement gap in year 2002-03 (see panel 2), the segregation index (H), pupil teacher ratio, and percentage of special educated students became the significant contributors to the variability in the district achievement gap in year 2002-03. The more segregated a district was, meaning that Hispanic students in the district tended to disproportionately cluster in some schools, the larger the achievement gap the district tended to have. Likewise, when pupil teacher ratio was higher in a district, or there were more special educated students, the achievement gap tended to be larger. The two variables “ELL student percentage” and “total student number” were not significant contributors to the variability in the district average achievement gap.

Panel 3 was about the district average growth rate. The segregation index H was found significantly related with the average growth rate, meaning that the more segregated a district was, the higher the average growth rate the district had. Percentage of special educated students and ELL students were also significantly related with district average growth rate, indicating that the more academically disadvantaged students in a district, the higher the growth rate the district tended to have.
Table 4.2.12
Relationship between District Variables and Math Achievement (Three-Level Conditional Model)

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>s.e.</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>For Intercept</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For Intercept (Initial status)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, $r_{000}$</td>
<td>488.4</td>
<td>2.5</td>
<td>.000</td>
</tr>
<tr>
<td>H (segregation index), $r_{001}$</td>
<td>-153.6</td>
<td>95.3</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Pupil teacher ratio, $r_{002}$</td>
<td>1587.8</td>
<td>523.5</td>
<td>.017</td>
</tr>
<tr>
<td>Special_ed students (%), $r_{003}$</td>
<td>-512.2</td>
<td>255.2</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>ELL students (%), $r_{004}$</td>
<td>-164.8</td>
<td>34.0</td>
<td>.001</td>
</tr>
<tr>
<td>Total student number, $r_{005}$</td>
<td>.0003</td>
<td>.0002</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>For Ethnicity (achievement gap)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, $r_{010}$</td>
<td>47.1</td>
<td>2.1</td>
<td>.000</td>
</tr>
<tr>
<td>H (segregation index), $r_{011}$</td>
<td>475.9</td>
<td>81.6</td>
<td>.000</td>
</tr>
<tr>
<td>Pupil teacher ratio, $r_{012}$</td>
<td>1445.4</td>
<td>458.7</td>
<td>.015</td>
</tr>
<tr>
<td>Special_ed students (%), $r_{013}$</td>
<td>600.9</td>
<td>204.1</td>
<td>.019</td>
</tr>
<tr>
<td>ELL students (%), $r_{014}$</td>
<td>53.3</td>
<td>29.0</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Total student number, $r_{015}$</td>
<td>0.2</td>
<td>0.2</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>For Time Slope</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For Intercept (Growth rate)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $r_{100}$</td>
<td>26.2</td>
<td>0.8</td>
<td>.000</td>
</tr>
<tr>
<td>H (segregation index), $r_{101}$</td>
<td>89.4</td>
<td>35.9</td>
<td>.037</td>
</tr>
<tr>
<td>Pupil teacher ratio, $r_{102}$</td>
<td>-293.3</td>
<td>178.3</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Special_ed students (%), $r_{103}$</td>
<td>211.5</td>
<td>83.8</td>
<td>.036</td>
</tr>
<tr>
<td>ELL students (%), $r_{104}$</td>
<td>35.0</td>
<td>10.9</td>
<td>.014</td>
</tr>
<tr>
<td>Total student number, $r_{105}$</td>
<td>-0.1</td>
<td>.06</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>For Ethnicity (achievement growth gap)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $r_{110}$</td>
<td>-5.1</td>
<td>0.6</td>
<td>.000</td>
</tr>
<tr>
<td>H (segregation index), $r_{111}$</td>
<td>-83.9</td>
<td>23.4</td>
<td>.008</td>
</tr>
<tr>
<td>Pupil teacher ratio, $r_{112}$</td>
<td>-230.0</td>
<td>131.2</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Special_ed students (%), $r_{113}$</td>
<td>-67.6</td>
<td>57.0</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>ELL students (%), $r_{114}$</td>
<td>-7.5</td>
<td>8.1</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Total student number, $r_{115}$</td>
<td>-0.03</td>
<td>.048</td>
<td>&gt;.05</td>
</tr>
</tbody>
</table>
The last panel showed the fixed effects on the district average growth gap. Among the five variables, only the segregation index was found to be significantly related. This suggests that the more segregated a district was, the larger were the average ethnic difference in growth rate between Hispanic and White students.

The random effects for school districts were displayed in the table below. The variance was decomposed to investigate whether residual variance still remains to be explained. The table confirmed that there were significant random effects at each level of the organizational units (student at level two and district at level three) for the average achievement and the average growth rate. With regard to the average achievement gap and the average growth gap, however, there was no significant variation at the district level. It implied that the ethnic difference in achievement gap and achievement growth gap did not vary much from district to district. However, this could also due to one limitation of the analysis that the district sample size was very small (N=9).

Table 4.2.13

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Variance Component</th>
<th>df</th>
<th>$\chi^2$</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporary variation, $e_{ij}$</td>
<td>456.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Level 2</strong> (students within districts)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual initial status, $r_{0ij}$</td>
<td>4129.2</td>
<td>13377</td>
<td>139042.6</td>
<td>.000</td>
</tr>
<tr>
<td>Individual growth rate, $r_{1ij}$</td>
<td>68.9</td>
<td>13377</td>
<td>17310.8</td>
<td>.000</td>
</tr>
<tr>
<td><strong>Level 3</strong> (between districts)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District average initial status, $\mu_{00j}$</td>
<td>47.3</td>
<td>12</td>
<td>38.2</td>
<td>.000</td>
</tr>
<tr>
<td>District average growth rate, $\mu_{10j}$</td>
<td>6.1</td>
<td>12</td>
<td>60.1</td>
<td>.000</td>
</tr>
<tr>
<td>District average initial gap, $\mu_{01j}$</td>
<td>12.3</td>
<td>12</td>
<td>13.7</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>District average growth rate gap, $\mu_{11j}$</td>
<td>1.2</td>
<td>12</td>
<td>13.8</td>
<td>&gt;.05</td>
</tr>
</tbody>
</table>
Summary

Different three level HLM models were employed to investigate the organizational effects of schools and districts. The purpose of investigating organizational effects is both to demystify the variability of achievement growth and achievement growth gap, and to link, schools and districts with the variability. The three level HLM models showed that schools varied from each other as to the four aspects - their initial achievement, initial achievement gap, achievement growth rate, and achievement gap in growth rate. The random effects of districts were only significant in average initial status and average growth rate. One possible explanation was that the number of districts may be too small to detect the significant variability across districts. It could also due to the fact that there was more variation in both initial status and growth rate at the school level than at the district level.

Among the five school variables in this study, the percentage of FRL students was found significantly related with the ethnic difference in achievement growth rate (achievement growth gap) while it was not a significant contributor to any other three aspects interested in the study (initial achievement, initial achievement gap, and achievement growth rate). The more low socio-economic students in a school, the larger difference in growth rate between Hispanic students and White students a school tended to have. It was also found that the more Hispanic students a school had, the more alike the students in the school tended to be , and the smaller the growth rate gap was.

Among the five district variables in this study, the segregation index (H) was the only one that was significantly related with the ethnic differences in growth rate between
Hispanic students and White students. The more segregated a district was, meaning that Hispanic students in the district disproportionately clustered in some schools, the smaller the achievement growth gap the district tended to have.

**Black-White Gap in Achievement Growth**

**Two-Level Conditional Model**

As for the Hispanic-White gap, four HLM conditional models are applied to investigate the Black-White achievement gap and growth gap, as well as to examine the relationship between initial score and growth rate.

**Model 1 (Two-Level Conditional Model with the Effect of Ethnicity)**

In model 1, the dummy variable of ethnicity (Black=0 White=1) is used to investigate the ethnic differences in student achievement and achievement growth. The table below compares the estimated fixed effects between the two-level conditional model for the Hispanic-White (H-W) gap and the one for the Black-White (B-W) gap.

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>H-W 11 p-value</th>
<th>B-W 12 p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for initial status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, $\beta_{00}$</td>
<td>485.2</td>
<td>502.7</td>
</tr>
<tr>
<td>Ethnicity contrast, $\beta_{01}$</td>
<td>58.4</td>
<td>41.2</td>
</tr>
<tr>
<td>Model for growth rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, $\beta_{10}$</td>
<td>26.2</td>
<td>22.9</td>
</tr>
<tr>
<td>Ethnicity contrast, $\beta_{11}$</td>
<td>-7.0</td>
<td>-3.7</td>
</tr>
</tbody>
</table>

11 H-W: The Hispanic-White gap  
12 B-W: The Black-White gap
Similar to the model for the Hispanic-White gap, the coefficients of ethnicity ($\beta_{01}$ 
\& $\beta_{11}$) were found significantly related to math initial achievement and achievement 
growth ($p<.01$). On average, White students started 41.2 points higher than Black 
students ($\beta_{01}$). The initial achievement gap between Black students and White students 
(41.2 points) was narrower than the one between Hispanic students and White students 
(58.4 points).

As to the individual growth rate, the scores of White students, on average, 
increased at a slower rate compared with their Black peers ($\beta_{11}$). Thus, when the scores 
of Black students increased, on average, at a rate of 22.9 points per year, the scores of 
White students increased, on average, at a rate of 19.2 ($= 22.9-3.8$) points per year. The 
Black-White gap in growth rate was also narrower than the Hispanic-White one.

In sum, the ethnic differences in both growth intercept and growth rate were 
significant. The results showed that achievement gap and growth gap existed between 
Black and White students in terms of their math achievement.

As to the random effect, the estimates for the variance of individual growth 
intercept $r_{0i}$ and growth rate $r_{1i}$ were both significant, indicating that individuals still 
varied significantly with regard to their initial scores and rates of change after their 
ethnicity were held constant.

As to the correlation between initial status and growth rate, the negative 
relationship was found between initial score and growth rate. The strength of correlation 
$\rho$ (- 0.70) was very close to the one for Hispanic and White students (-0.67).
Model 2 (Holding Initial Status Constant)

The strong negative correlation between growth intercept and growth rate provides a great rationale that growth intercept should be controlled for when we compare the growth rates between Black students and White students. The table below presents the results of fixed effects of conditional model with initial status controlled for.

Table 4.2.15
Fixed Effects of the Two-Level Conditional Model with Initial Status Controlled for

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>H-W</th>
<th>p value</th>
<th>B-W</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for initial status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, $\beta_{00}$</td>
<td>485.2</td>
<td>.000</td>
<td>502.7</td>
<td>.000</td>
</tr>
<tr>
<td>Ethnicity contrast, $\beta_{01}$</td>
<td>58.7</td>
<td>.000</td>
<td>41.2</td>
<td>.000</td>
</tr>
<tr>
<td>Model for growth rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, $\beta_{10}$</td>
<td>22.1</td>
<td>.000</td>
<td>19.4</td>
<td>.000</td>
</tr>
<tr>
<td>Ethnicity contrast, $\beta_{11}$</td>
<td>-1.9</td>
<td>.000</td>
<td>-0.01</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Initial score, $\beta_{12}$</td>
<td>-0.1</td>
<td>.000</td>
<td>-0.1</td>
<td>.000</td>
</tr>
</tbody>
</table>

The table above showed that after initial scores were controlled for, the gap in growth rate ($\beta_{11}$) was not significantly related to math achievement ($p > .05$), while gap in growth intercept still existed ($\beta_{01}$). On average, White students still started 41.2 points higher than Hispanic students ($\beta_{01}$). As to the individual growth rate, no significant ethnic difference was found between Black students and White students. This was different from the Hispanic-White growth gap, where the test scores of White students increased, on average, at a rate of 1.9 points per year slower than their Hispanic peers ($\beta_{11}$). Therefore, when initial scores were controlled for, the Black-White gap in growth rate was not significant any more. Stated differently, if White students and Black students
started with the same scores, White students would grow as fast as Black students. To decrease the achievement gap, it is hoped that the gap in growth rate can be large so that minority students can grow faster than their White peers over the time. However, the result here suggests that after the initial status being controlled for, there was no growth gap between the two ethnic groups.

The coefficient of initial status $\beta_{12} (-0.1)$ indicates that after student ethnicity being held constant, when initial score increases one point, growth rate would decrease 0.1 point. For example, if two students are of the same ethnicity, when student A’s initial score is 10 points higher than student B, student A is expected to grow 1 point slower than student B. However, the model cannot tell whether the correlation between initial status and growth rate is constant across ethnic groups. Stated differently, it is of concern whether Black students and White students differed from each other significantly as to the relationship between the two growth parameters. Model 3 is applied then when the interaction effect between ethnicity and initial status being held constant at the level-two model for growth rate.

Another notable feature about this model was that just as was the case with the model for the Hispanic-White gap, there was no significant variance for the slope of individual growth rate. It indicates that growth rate did not vary significantly from individual to individual while initial status was controlled for.

**Model 3 (Holding the Interaction Effect between Ethnicity and Initial Status Constant)**

In order to control for the interaction between ethnicity and initial status, the interaction effect is included into the second level model for growth rate.
Based on Table 4.2.16, the results for initial status remain the same. On average, White students started 41.2 points higher than Black students (\( \beta_{01} \)). As to individual growth rate (\( \beta_{11} \)), the ethnic difference was still *not significant* after the interaction effect was controlled for. The interaction effect of ethnicity with initial status \( \beta_{13} \) was found *not significant* as well. It means that the relationship between initial score and growth rate was constant across different ethnic groups. For both Hispanic students and White students, when their scores started low, they grew more; and when they started high, they grew less.

Compared with model 2, the prediction of initial status on growth rate remains significant, suggesting that initial status was a more important predictor than ethnicity. The result showed that as to growth rates, whether students are Black or White did not matter as much as where their initial scores were.

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>H-W</th>
<th>p-value</th>
<th>B-W</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for initial status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, ( \beta_{00} )</td>
<td>485.2</td>
<td>.000</td>
<td>502.7</td>
<td>.000</td>
</tr>
<tr>
<td>Ethnicity contrast, ( \beta_{01} )</td>
<td>58.7</td>
<td>.000</td>
<td>41.2</td>
<td>.000</td>
</tr>
<tr>
<td>Model for growth rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, ( \beta_{10} )</td>
<td>21.5</td>
<td>.000</td>
<td>16.9</td>
<td>.000</td>
</tr>
<tr>
<td>Ethnicity contrast, ( \beta_{11} )</td>
<td>-1.0</td>
<td>&gt;.05</td>
<td>2.6</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Initial score, ( \beta_{12} )</td>
<td>-0.1</td>
<td>.000</td>
<td>-.09</td>
<td>.000</td>
</tr>
<tr>
<td>Ethnicity x Initial scores ( \beta_{13} )</td>
<td>-0.002</td>
<td>&gt;.05</td>
<td>-.005</td>
<td>&gt;.05</td>
</tr>
</tbody>
</table>
Model 4 (Holding Other Student-Level Variables Constant)

Other student variables such as gender (Male=0, Female=1) and student mobility were also investigated to see whether they were significantly correlated with initial status and growth rate. The table below displays the fixed effects of the two-level conditional model with person-level predictors.

Table 4.2.17

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>H-W</th>
<th>p-value</th>
<th>B-W</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model for initial status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, $\beta_{00}$</td>
<td>486.8</td>
<td>.000</td>
<td>505.0</td>
<td>.000</td>
</tr>
<tr>
<td>Ethnicity contrast, $\beta_{01}$</td>
<td>64.8</td>
<td>.000</td>
<td>46.6</td>
<td>.000</td>
</tr>
<tr>
<td>Gender (Male), $\beta_{02}$</td>
<td>4.2</td>
<td>&gt;.05</td>
<td>5.6</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Mobility, $\beta_{03}$</td>
<td>-6.3</td>
<td>.000</td>
<td>-8.7</td>
<td>.019</td>
</tr>
<tr>
<td>Ethnicity x Gender, $\beta_{04}$</td>
<td>-7.3</td>
<td>.015</td>
<td>-8.6</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Ethnicity x Mobility, $\beta_{05}$</td>
<td>-9.8</td>
<td>.000</td>
<td>-7.5</td>
<td>.049</td>
</tr>
<tr>
<td><strong>Model for growth rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, $\beta_{10}$</td>
<td>23.0</td>
<td>.000</td>
<td>15.6</td>
<td>.026</td>
</tr>
<tr>
<td>Ethnicity contrast, $\beta_{11}$</td>
<td>-2.9</td>
<td>&gt;.05</td>
<td>3.6</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Gender (Male), $\beta_{12}$</td>
<td>-1.8</td>
<td>.008</td>
<td>1.6</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Mobility, $\beta_{13}$</td>
<td>-1.0</td>
<td>.003</td>
<td>0.4</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Initial score, $\beta_{14}$</td>
<td>-0.9</td>
<td>.000</td>
<td>-0.8</td>
<td>.000</td>
</tr>
<tr>
<td>Ethnicity x Initial scores, $\beta_{15}$</td>
<td>-0.002</td>
<td>&gt;.05</td>
<td>-0.006</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Ethnicity x Gender, $\beta_{16}$</td>
<td>2.3</td>
<td>.002</td>
<td>-1.0</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Ethnicity x Mobility, $\beta_{17}$</td>
<td>1.2</td>
<td>.002</td>
<td>-0.2</td>
<td>&gt;.05</td>
</tr>
</tbody>
</table>

The above table shows that at year 2002-03 White students, on average, scored 46.6 points ($\beta_{01}$) higher than their Black peers when student gender and mobility were held constant. Compared with model 3 (41.2 points), the addition of person-level predictors (gender and mobility) widened the ethnic gap in initial scores. Neither gender
difference nor interaction effect between gender and ethnicity was found significant. Students with different moving patterns did score differently and the difference was confounded with their ethnicity. Black students, no matter their moving pattern, tended to score lower than White peers.

As to the model for growth rate, ethnicity was found not significantly related with growth rate ($\beta_{11}$). There was no significant ethnic difference between Black students and White students in how fast their math scores increased across the three years, when initial status, the interaction of initial status with ethnicity, gender, and student mobility were held constant. The other student characteristics, such as gender and mobility, were not found significantly related with growth rate. In such a model for Hispanic and White students, not only initial status was found significantly related with growth rate, there was also significant gender and mobility difference in individual growth rate. The disparities between the two models indicate that for different gaps, there were different explanations and interpretations. The results further suggest it is necessary to display the scenarios for each type of achievement gap.

Three-Level Conditional Model (Level Three - School)

At level three, the variability among schools/districts in the growth parameters is addressed. The explanatory school variables include the minority (Black) percentage in a school, school mobility rate, total student number, pupil-teacher ratio, and percentage of students who had Free or Reduced-price Lunch (FRL). The table below displays the fixed effects of student ethnicity and school variables on math achievement.
Table 4.2.18  
*Relationship between School Variables and Math Achievement (Three-Level Conditional Model)*

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>H-W</th>
<th>p value</th>
<th>B-W</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>For Intercept</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For Intercept (initial status)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, $r_{000}$</td>
<td>496.7</td>
<td>.000</td>
<td>445.6#</td>
<td>.000</td>
</tr>
<tr>
<td>Minority percentage, $r_{001}$</td>
<td>-44.4</td>
<td>.008</td>
<td>255.6#</td>
<td>.005</td>
</tr>
<tr>
<td>Mobility rate, $r_{002}$</td>
<td>-100.5</td>
<td>.000</td>
<td>24.0</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Total student number, $r_{003}$</td>
<td>0.02</td>
<td>.004</td>
<td>-0.02</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Pupil teacher ratio, $r_{004}$</td>
<td>0.2</td>
<td>&gt;.05</td>
<td>3.5</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>FRL students (%), $r_{005}$</td>
<td>-28.5</td>
<td>&gt;.05</td>
<td>-297.0#</td>
<td>.002</td>
</tr>
<tr>
<td><strong>For Ethnicity (achievement gap)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, $r_{010}$</td>
<td>43.6</td>
<td>.000</td>
<td>171.2</td>
<td>.093</td>
</tr>
<tr>
<td>Minority percentage, $r_{011}$</td>
<td>12.3</td>
<td>&gt;.05</td>
<td>-295.8</td>
<td>.004</td>
</tr>
<tr>
<td>Mobility rate, $r_{012}$</td>
<td>-3.1</td>
<td>&gt;.05</td>
<td>-119.9</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Total student number, $r_{013}$</td>
<td>-0.01</td>
<td>&gt;.05</td>
<td>0.04</td>
<td>.005</td>
</tr>
<tr>
<td>Pupil teacher ratio, $r_{014}$</td>
<td>-1.4</td>
<td>&gt;.05</td>
<td>-6.9</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>FRL students (%), $r_{015}$</td>
<td>-34.7</td>
<td>&gt;.05</td>
<td>210.4</td>
<td>.027</td>
</tr>
<tr>
<td><strong>For Time Slope</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For Intercept (growth rate)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, $r_{100}$</td>
<td>25.6</td>
<td>.000</td>
<td>67.8</td>
<td>.005</td>
</tr>
<tr>
<td>Minority percentage, $r_{101}$</td>
<td>13.0</td>
<td>.014</td>
<td>-32.3</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Mobility rate, $r_{102}$</td>
<td>6.5</td>
<td>&gt;.05</td>
<td>21.6</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Total student number, $r_{103}$</td>
<td>-0.06</td>
<td>.002</td>
<td>-0.002</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Pupil teacher ratio, $r_{104}$</td>
<td>-0.067</td>
<td>&gt;.05</td>
<td>-2.3</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>FRL students (%), $r_{105}$</td>
<td>-5.5</td>
<td>&gt;.05</td>
<td>24.0</td>
<td>&gt;.05</td>
</tr>
<tr>
<td><strong>For Ethnicity (achievement growth gap)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base, $r_{110}$</td>
<td>-5.3</td>
<td>.000</td>
<td>-36.5</td>
<td>.146</td>
</tr>
<tr>
<td>Minority percentage, $r_{111}$</td>
<td>-6.9</td>
<td>.025</td>
<td>41.0</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Mobility rate, $r_{112}$</td>
<td>7.1</td>
<td>&gt;.05</td>
<td>9.4</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Total student number, $r_{113}$</td>
<td>.002</td>
<td>&gt;.05</td>
<td>-0.007</td>
<td>.044</td>
</tr>
<tr>
<td>Pupil teacher ratio, $r_{114}$</td>
<td>0.1</td>
<td>&gt;.05</td>
<td>1.8</td>
<td>&gt;.05</td>
</tr>
</tbody>
</table>
Starting from Panel 1, as to the school average achievement in year 2002-03, percentage of FRL students was found significant, indicating the higher percentage of FRL students in a school, the lower the school average achievement score. Minority percentage, although a significant predictor of the school average achievement scores both for the H-W model and for the B-W model, the direction of the relationship was opposite. In the model for the Black-White gap, the higher percentage of Black students in a school, the higher the school average score in year 2002-03. In the model for the Hispanic-White gap, the lower percentage of Hispanic students in a school, the higher the school average score in year 2002-03.

As for the school average achievement score gap in year 2002-03 (panel 2), when no school variable was found significantly related with the Hispanic-White gap, minority percentage, total student number (school size), and percentage of FRL students were significantly related with the Black-White gap. Hence, the lower the percentage of Black students, the larger the school size, or the higher percentage of FRL students in a school, the wider the achievement gap between Black and White students the school tended to have.

No school characteristic was found to be significantly related with the school average growth rate (see panel 3). For the schools which had Hispanic students and

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>H-W</th>
<th>p value</th>
<th>B-W</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRL students (%)</td>
<td>10.3</td>
<td>.007</td>
<td>-21.6</td>
<td>&gt;.05</td>
</tr>
</tbody>
</table>

*Note: The unusually large coefficients here may be due to sparse data available for the Black-White comparisons*
White students, things were different. The minority percentage and the total student number were the significant predictors of the school average growth rate.

The last panel in Table 4.2.18 represents the concept of ethnic differences in achievement growth in this study, which is also called achievement growth gap. Only school size was found significantly related with the school average growth gap. The larger a school was, the smaller the difference in growth rate between Black students and White students. Thus, Black students and White students tended to have more similar growth rates in the larger schools. It was also noticeable that 1) percentage of FRL students was significantly related with the growth gap between Hispanic and White students but not with the growth gap between Black and White students (the more FRL students, the larger the school average growth gap), and 2) minority percentage was significantly related with the Hispanic-White achievement growth gap but not with the Black-White growth gap (the larger the Hispanic percentage, the smaller the school average growth gap). The results showed that the achievement gaps should be analyzed and treated differently since different school characteristics may be related with different growth gaps.

In order to examine whether the residual variance of student math achievement still remains to be explained, the variance was decomposed into three levels. No significant random effects at level three was found. Stated differently, the schools, which had Black students, did not significantly vary from each other with regard to their average achievement in 2002-03, average achievement gap in 2002-03, average growth rate, and average achievement growth gap. This is not consistent with what was found for
the schools that had Hispanic and White students, where schools varied with respect to all these four aspects.

**Three-Level Conditional Model (Level Three - District)**

When the third-level units were replaced by districts, the district characteristics on student achievement and achievement growth were then investigated. Among the five predictors - segregation index (H), pupil teacher ratio, percentage of special educated students, percentage of FRL students, and number of students, none of the district characteristics was found significantly related with the district average scores, district average Black-White gap, district average growth rate, and district average Black-White growth gap.

Table 4.2.19  
*Variance Decomposition from a Three-Level Analysis (Level 3 - District)*

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Variance Component</th>
<th>p value (H-W)</th>
<th>Variance Component</th>
<th>p value (B-W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporary variation, ( e_{ij} )</td>
<td>456.8</td>
<td></td>
<td>434.9</td>
<td></td>
</tr>
<tr>
<td>Level 2 (students within districts)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual initial status, ( r_{0ij} )</td>
<td>4129.2</td>
<td>.000</td>
<td>4151.1</td>
<td>.000</td>
</tr>
<tr>
<td>Individual growth rate, ( r_{ij} )</td>
<td>68.9</td>
<td>.000</td>
<td>65.7</td>
<td>.000</td>
</tr>
<tr>
<td>Level 3 (between districts)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District average initial status, ( \mu_{00j} )</td>
<td>47.3</td>
<td>.000</td>
<td>75.4</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>District average growth rate, ( \mu_{10j} )</td>
<td>6.1</td>
<td>.000</td>
<td>0.6</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>District average initial gap, ( \mu_{01j} )</td>
<td>12.3</td>
<td>&gt;.05</td>
<td>143.8</td>
<td>.001</td>
</tr>
<tr>
<td>District average growth rate gap, ( \mu_{11j} )</td>
<td>1.2</td>
<td>&gt;.05</td>
<td>10.1</td>
<td>.022</td>
</tr>
</tbody>
</table>

The random effects for school districts are displayed in the table below. The variance was decomposed to investigate whether residual variance still remains to be
explained. The table confirms that there were significant random effects at student level for the average achievement and the average growth rate. At the district level, significant random variation was only found in district achievement gap and district growth gap, indicating that districts were significantly different from each other in the ethnic differences in initial scores and growth rates between Black students and White students. It was different from the model for the Hispanic-White gap, where random variability was only significant for district average initial scores and district average growth rates. It further suggested that there were two quite different scenarios for these two gaps.

**Summary**

Similarities and differences were found in the findings in the investigations of the Black-White gap and the Hispanic-White gap. First, the Black-White gap was narrower than the Hispanic-White one in terms of the initial status and the growth rate. Second, initial score was a very important predictor of growth rate in both gaps. After initial status was controlled for, the Black-White gap in growth rate became no longer significant. Where students started, rather than their ethnic background, played a critical role in predicting how fast students grew. Third, similar to the Hispanic-White model, the interaction effect was not significant in predicting the growth rate.

At the school level, different school characteristics were related with different gaps, suggesting that separate analyses are always necessary for different gaps. At the district level, no district characteristics was found significantly related with the district average scores, district average Black-White initial gap, district average growth rate, and district average Black-White growth gap. A possible explanation is that since the number
of districts was small (N=12), the power of the analysis may be inadequate. It is also possible that the variation in the district variables was not substantial enough.

As to the random effects of schools, when the three level HLM models for Hispanic and White students showed that schools varied from each other with respect to the four interesting aspects—school average initial achievement, school average initial achievement gap, school average achievement growth rate, and school average gap in growth rate, no significant school random effects for Black and White students was found. Stated differently, the schools with Black students did not significantly vary from each other as to their average achievement and achievement gaps. This could be because most of the schools contained a very small proportion of Black students, the variation of the schools was not significant.

At the district level, significant random variation was only found in district average achievement Black-White gap and district average Black-White growth gap, indicating that districts were significantly different from each other in the *ethnic differences* in initial scores and growth rates. It was different from the random variability for the Hispanic-White model, where random variability was only significant in terms of district average initial scores and district average growth rates. The results further suggest that the districts with Black and White students somehow had similar growth intercepts and rates, but the score differences as well as the differences in growth rate between the two groups varied from district to district. On the other hand, for the districts with Hispanic and White students, the score differences and growth rate differences between
the two groups did not vary from district to district, but the initial scores and growth rates of these districts varied significantly from each other.

The comparisons between the two gaps further validate our conclusion that different descriptions, interpretations, and explanations should be employed for analyzing different gaps.
Chapter 5: Discussion

Returning to our initial research question of whether ethnicity is a meaningful indicator of educational inequality, the research conveys the message that even if student ethnicity is confounded with other indicators, such as initial score and socio-economic status, it is still a very important predictor of both achievement gaps and achievement growth gaps.

NCLB was intended to ensure that all schools set high standards for reading and math, and to hold all students accountable to these standards, regardless of race, income, or other differences. The achievement gap between ethnic groups has been one of the major concerns in discussions of educational inequalities. Different reports have showed that the achievement gap between European American (White) students and African American (Black) students was decreasing since the Civil Rights movement although the rate of decrease slowed since 1990’s. The initial research question for this study was whether the achievement gaps between minority students (African American and Hispanic American) and non-minority students (European American) have decreased after the passage of NCLB. To answer this question, the traditional descriptive approach displayed and compared scale scores or percentages above proficient among different groups. We adopted this approach in the dissertation by comparing mean scores and effect sizes instead of percent above proficient. In addition, we treated achievement gaps as ethnic differences in achievement growth (achievement growth gap) rather than static differences at multiple time points. The study further explored what factors were associated with the achievement growth gaps. The residual variance of student math achievement was disaggregated into the individual, school and district levels, and was associated with the four aspects of interest: 1) initial achievement (achievement scores in
year 2002-03), 2) initial achievement gap (ethnic difference in achievement scores in year 2002-03), 3) achievement growth (growth rate over the three years), and 4) achievement growth gap (ethnic difference in growth rate over the three years).

As an attempt to explore new applications of HLM, this study investigated the achievement gap using the HLM growth modeling approach. With the advantage of tracking one cohort of students, this approach helped explain the new concept of *ethnic differences in achievement growth (achievement growth gap)*. Moreover, the model addressed the relationship between initial status and rate of change, the relations of person-level and school/district-level predictors to both status and change, and the random effects of students, schools, and districts.

Both descriptive analyses and HLM growth modeling were applied to examine trends in the achievement gaps after the passage of NCLB. The two approaches are compared to address the question of whether methodology matters in terms of understanding achievement gaps. The comparison reveals similar trends in the achievement gaps when disparities exist. The growth modeling provides a fresh perspective on the issue of ethnic inequality in an educational system, and the new term *ethnic difference in achievement growth* was illustrated to support the consideration of potential policy changes. Limitations of the study are reviewed to ensure caution in the generalization of the results.

**Comparison between the Descriptive Analyses and the HLM Growth Modeling**

When investigating student achievement and school effectiveness, only a few researchers have explored the impact of methodological strategies on the analysis of
educational policy. One example was that Raudenbush in 2004 discussed the differences in the results of accountability systems based on student mean proficiency and those based on value added modeling. Although the results were mostly similar, Raudenbush found that measures based on mean proficiency “scientifically indefensible for high-stake decisions.” In our study, the comparison between the two approaches aimed to provide a comprehensive view of achievement gaps. Although the growth modeling did not give substantially different results as to the general trend of the achievement gaps, differences between the two approaches existed in many details which were associated with the methodological differences. The HLM approach especially expanded our understanding of achievement gaps by addressing 1) the new concept of ethnic difference in achievement growth, 2) the relationship between where students start and how fast they grow, and 3) the random effects and the fixed effects at different levels.

*The Concept of Ethnic Differences in Achievement Growth (Achievement Growth Gap)*

The term achievement gap is often applied in cross-sectional studies in which the achievement of different cohorts of students is compared over years. In this study, both scale scores and effect sizes were employed to depict the achievement gaps among different ethnic groups based on a single cohort of students. The scale score gaps between minority and non-minority students were calculated over the three academic years. On average, Hispanic students scored lowest and White students scored highest. The gap between White and Hispanic students narrowed gradually across the years while the gap between White and Black students only decreased slightly. The perspective afforded by the use of effect sizes revealed a slightly different trend: The standardized difference in
the mean scale scores between Hispanic and White students slightly decreased while the standardized difference between Black and White students did not really decrease.

The HLMs represent how the achievement scores of different ethnic groups grew over years. Conceptually, the specified growth model can be viewed as a within-person regression model representing individual change over time. The term *ethnic difference in achievement growth* refers to the difference in achievement growth intercepts and growth rates between minority and non-minority students based on the observations of the same cohort of students at multiple time points. The growth intercept, indicating the student initial achievement score in year 2002-03, and the growth rate, indicating the rate of change of the achievement score every year, are the focus of the study.

Similarities and disparities in the results between the two approaches are given below:

i) Both approaches found that Hispanic students, on average, had the lowest initial scores and White students’ average initial score was the highest. The achievement score gap in the starting year between Hispanic students and White students was larger than the one between Black students and White students.

ii) Both approaches found that Hispanic students grew most either in terms of their mean gain score or mean growth rate and White students grew least. The achievement growth gap between Hispanic students and White students was larger than the one between Black students and White students. Although the general trend in growth was the same, the gain score and the growth rate conveyed different ideas of growth. The mean gain score here was the difference
between the mean last-year score and the mean first-year score. It basically ignored the information of the scores in between the two data points. HLM, on the other hand, captured the full length of the achievement growth since the growth rate is based on all the data points. The more data points, the more precise the estimation.

iii) Both approaches found the gaps were decreasing. The score gaps between ethnic groups were decreasing. As to the growth modeling, Minority students were found to grow faster than their White peers.

iv) Both approaches showed that taking into account the variability of the scores provided an insight into the investigation of achievement gap. The effect size of the score gaps showed that the Black-White standardized mean difference was almost constant over the three years. The HLM approach, however, found that Black students grew significantly faster than their White peers. The two-level model also demonstrated that individual students varied considerably with respect to their growth rates, suggesting the need for additional modeling. To attempt to account for variability at the individual level, initial status was included in the model. The residual variation in individual growth rates became insignificant and the ethnic difference between Black students and White students in growth rates disappeared. Hence, although both approaches can address the issue of variability of scores, effect size can only characterize differences at the aggregate level, while with HLM it is possible to examine the individual variability. More importantly, the HLM approach can link the
individual residual variability with individual characteristics (covariates), and thus the residual variability can be further explained by the covariates included in the model.

In sum, the two approaches revealed similar general trends in achievement gaps. Beyond looking at the two data points to compute a gain score, HLM provided us with a growth trajectory by computing the mean growth rate across the three years. Knowing how fast students in each ethnic group grew, we can compare the growth rates. For example, Hispanic students and Black students, on average, grew 7.0 and 3.7 more points per year than their White peers. Based on the current growth intercepts and growth rates of each ethnic group, we can predict how long the achievement gap can be closed. For example, the Hispanic-White gap may be closed in around eight years with the initial score gap (58.4 points) and the current growth rate difference (7.0 points). The Black-White gap may be closed in around eleven years with the initial score gap (41.2 points) and the current rate difference (3.7 points). Both conclusions are under the assumption that the current growth rates will persist.

Another advantage of the HLM technique is that variability of individual scores can be taken into account in the model. Instead of using the total variance in effect size, the growth model disaggregates the score variability into within and between individual random variations. The HLM growth models for both the Hispanic-White gap and the Black-White gap revealed that there was a significant amount of variation in growth intercept and in growth rate at the individual level. After initial status was controlled for, the variance in growth rate across individuals was no longer significantly different from
zero. That is, if all students started at the same scores, then their individual scores would be expected to increase at similar rates. This hypothetical circumstance, however, is not the case for this cohort. Students did start from different scores, and minority students, on average, had lower initial scores.

In sum, the new term *ethnic difference in achievement growth* effectively captured the trend of the achievement gap. The concept can be used to predict the achievement gap and to control for the effect of initial status when the growth intercepts and growth rates between ethnic groups are compared.

**The Relationship between Where Student Start and How Fast They Grow**

One very important advantage of applying HLM to the study of achievement gaps is that the relationship between where students start and how fast they grow can be investigated. The descriptive approach, by comparing the mean scores of different groups, hinted that the starting score might be related with the gain score. For example, Hispanic students had the lowest average starting score but the highest average gain score. The results of the HLM approach confirmed that initial score was a significant predictor of growth rate. Note that the notion of starting score and initial score are exchangeable since in this study they both refer to the achievement scores in year 2002-03.

Both approaches found that minority students, on average, started with lower scores but ended with larger gain scores (or higher growth rates). Yet differences existed. When starting scores were controlled for, the ethnic differences in gain score were still significant. It indicates that even if students all had started at the same scores, minority
students would have still gained more than non-minority students. In the HLM growth models, with initial status being controlled for, the ethnic difference in growth rate between Black students and White students was not significant any more, and the gap in growth rate between Hispanic and White students became narrower (1.9 points per year) although it was still significant. It suggests that if minority students and non-minority students had started at the same scores, Blacks students and White students would have grown at similar rates while Hispanic students would have grown 1.9 points per year faster than their White peers. Under this circumstance, one can argue that although with the initial status controlled for, Black students’ average growth rate was not significantly higher their White peers. In order to narrow the achievement gap, Black students would have to have a higher growth rate so that the score differences between them and White students would narrow over time.

The relationship between where students start and how fast they grow was further examined by controlling for the interaction between ethnicity and starting score. After the interaction is controlled for, the growth modeling showed that the differences in growth rates between Hispanic and White students, as well as between Black and White students, were not significant. The traditional ANOVA approach, however, displayed another picture: There was a significant ethnic difference in gain scores after the interaction between starting scores and gain scores is controlled for. It indicates that ethnic differences in gain scores changed with the location of initial scores. For example, at the low end of the range of initial scores, the mean gain score of Hispanic students was significantly higher than the mean gain scores of the other two groups. At the middle and
the high end of the range, however, the mean gain score of Hispanic students turned to be lower than that of White students, although the difference was not significant.

The ANOVA results about gain score were different from what was found about growth rate. The ethnic differences in growth rates did not vary significantly at different locations of initial scores. For example, Hispanic students grew faster than their White peers at both the low end and the high end of the range of initial scores. The results of growth modeling were more stable since for different locations of initial scores, the ethnic gaps in growth rate remained constant. Instead of the achievement scores at one-time or multiple-time points, the focus of the HLM growth modeling was on the growth rates, the growth intercepts, and the relationship between the two growth parameters. While the descriptive approach focused on the achievement gaps in gain score, the growth modeling provided a relatively consistent view on the projection of student growth.

**Random Effects and Fixed Effects at Different Levels**

A salient feature of HLM is that it recognizes the hierarchical structure of the educational system and represents the random variability across students, schools and districts. The characteristics of the organizational units in the hierarchical structure are treated as *fixed effects* -- these are the covariates at each level in the school system. In HLM, they are used to account for some of the variability at the individual, school and district levels (*random effects*) to provide a full representation of the patterns in student achievement and achievement gaps.
The descriptive approach to analyzing achievement gaps compared the mean scale scores and the effect sizes among ethnic groups. The comparison of the mean outcomes, to a great extent, neglected the real differences among different organizational units. For example, when we compared the scale scores at the individual, school and district levels, it showed that no matter what ethnic group students belonged to, they tended to score similarly if they were in the same schools and districts. The differences among schools and districts were not taken into consideration by the descriptive approach. Such a simple analysis can lead to an incomplete or even invalid conclusion. Meanwhile, for the descriptive analyses, there is always an implied assumption of the homogeneity of variance, which is not the case in our study.

At the individual level, both approaches found that female students, on average, scored slightly lower than their male peers but had higher average gain scores (or grew faster). The gender differences were not statistically significant except that Hispanic female students had a significantly higher average growth rate than Hispanic male students. Both approaches found that, on average, students who moved scored higher than those who did not move, that students who moved due to bureaucratic reasons scored higher than those who moved for other reasons, and that students who moved once scored higher than those who moved twice. Analyses based on the HLM models also found that for White students and Hispanic students, students with less mobility had higher average growth rates. The significant correlation between mobility and growth rates was not found for the group of Black students and White students. As to the random effects of the two-level HLM models, students differed from each other significantly in
their initial scores and growth rates. After the initial status was held constant for the group of Hispanic and White students, the significant variation in individual growth rates became non-significant, suggesting that students would not differ significantly in their individual growth rates if they started at a same initial score.

Employing three-level HLM models, we examined how various school/district characteristics were related to the variability in the four key parameters at the school or district level: average initial status, average initial gap, average growth rate, and average growth gap. To investigate the relationship between school / district characteristics and the four aspects of interest, correlational analyses were conducted with variables of interest linked separately. By contrast, the HLM models connected the variations in these four aspects with school and district characteristics and investigated how they interacted as a system.

Similarities and differences were found between the traditional analytic approach and the HLM analyses. First, both approaches found that students tended to score similarly if they were in same schools and districts, especially in terms of how fast they grow. The percent of total variance in initial status that lies between schools and between districts is 19% and 9%, respectively. A larger proportion of the variance in growth rates lies between schools and between districts (36% and 20%, respectively). A large amount of variance (81%) in growth intercept (initial status) can only be explained at the individual level while almost one-third of variance in growth rate can be explained at the school level. Schools mattered more than districts as to how fast students’ scores were growing.
Second, differences between the two approaches were found as to which school characteristics were significantly correlated with school average initial achievement gap and school average gap in growth rate. Among the five selected school variables—minority percentage, school mobility rate, total student number, pupil teacher ratio, percentage of FRL students, both approaches found that no school variable significantly correlated with the school average initial gap between Hispanic students and White students. For the initial gap between Black students and White students, no significant correlation was found in descriptive correlation analyses. In HLM, minority percentage, total student number, and FRL student percentage were found to be significantly related with the initial Black-White (B-W) score gap at the school level. This might be because the achievement scores of Black students were highly variable, and when the random variation of scores of Black students were controlled for in the HLM models, fixed effects were more precisely estimated. Similar disparities were found in the analysis of school average gap in growth. In descriptive analyses, no school variable was found to significantly correlate with the gain score gap. In HLM, minority percentage and FRL student percentage were significantly correlated with the Hispanic-White (H-W) growth gap, while total student number was significantly correlated with the B-W growth gap.

Third, at the district level, among the five district variables—segregation index (H), pupil teacher ratio, percentage of special educated students, percentage of FRL students, and total student number, both approaches found that the segregation index was a significant predictor to the initial achievement gap and the gap in growth (or gain score) between Hispanic students and White students. In HLM, it was also found that pupil
teacher ratio was a significant predictor of the H-W initial achievement gap. Neither of the two approaches found that any district characteristic was significantly related with the district average Black-White achievement gap and the district average Black-White gap in growth rate.

Lastly, the random variation at the school and district levels can only be examined by the HLM approach. In order to investigate whether residual variance of student math achievement still remained to be explained, the variance was decomposed into three levels. For the schools that had Hispanic and White students, after controlling for some school characteristics, a significant amount of residual variance was still found at the school level. The results suggested that the current model could not fully explain the variance among schools and some other possible school characteristics should be included in the future study to further explain the significant variability. For the schools that had Black and White students, no significant random variability at the school level was found. The results suggested that for this sample of schools, most variation among schools can be explained by the current model.

At the district level, in the model for the Hispanic-White gap, significant random variation was only found in the district average achievement gap and the district average growth gap, indicating that districts were significantly different from each other in the ethnic differences in initial scores and growth rates between Hispanic students and White students. This was different from the model for the Black-White gap, where random variability was only significant in district average initial scores and district average growth rates. The results indicated that as to the gaps and the growth gaps between
Hispanic and White students, districts differed significantly from each other; but they were somehow homogeneous with regard to the gaps between Black and White students. The disparities in results between the Hispanic-White gaps and the Black-White gaps suggest that it is necessary to analyze the two sets of gaps separately.

**Policy Implications**

The school accountability system based on *status measures* has been criticized for being inappropriate for evaluating educational effectiveness because schools serving academically disadvantaged students tend to be judged unfairly. More and more states are now incorporating growth modeling into their school accountability system to recognize “underperforming” schools whose students are making progress. Similarly, the concept of ethnic differences in achievement growth, which is also called the achievement growth gap, examines achievement gaps from a growth perspective. Starting from the question of whether ethnicity is still a meaningful indicator of educational inequality, the study found that the ethnic gaps in *achievement scores* still existed. The concept of ethnic differences in *achievement growth* was introduced to determine whether there were gaps in how fast different ethnic groups of students grew and whether achievement gaps were narrowing. The new definition aims to help policymakers and other stakeholders better understand the issue of the achievement gaps so that they can formulate more effective policies and strategies. For example, with the knowledge of growth intercept and growth rate, we are not only able to predict when gaps will be closed if current rates persist, but also can assist in detecting the effectiveness of interventions by tracking changes in growth rates.
It is commonly understood that student initial scores may correlate with their later achievement. Not much research, however, has been done as to the relationship between where students start (initial status) and how fast they grow (growth). This study has shown that understanding of both growth and growth gaps will be incomplete or even misleading if we only compare average growth rates between ethnic groups. In this study, it was found that although the scores of Black students and Hispanic students increased at a significantly higher rate than their White peers, holding their initial status constant effectively eliminated the difference in growth rates between Black students and White students. The gap in growth rates between Hispanic students and White students remained statistically significant but became much narrower. Therefore, the conclusion of whether the achievement gaps are decreasing should be made cautiously. On the one hand, the scores of minority students grew faster and the score gaps were decreasing. On the other hand, the estimation of growth rate was greatly confounded with student initial status. The initial test scores explained a significant amount of variation in the relationship between ethnicity and growth rate. Hence, the concept of ethnic difference in growth should always include growth intercept, growth rate and the relationship between the two. Only considering one aspect of growth could lead to invalid policy implications. For example, when an educational program is designed to close the achievement gap in a school district, ethnicity should not be the only indictor to report. Student initial status as well as its relationship with growth rate should be integrated into the evaluation of a program’s effectiveness.
Although individuals from different ethnic groups with the same initial scores may experience similar growth rates, empirically it is the case that Hispanic students and Black students usually score lower than their White peers. We expect minority students to grow at a faster rate and, hence, reduce the achievement gaps over time. Even though initial status can explain a significant amount of variance in growth rate, it does not mean that achievement gaps in growth do not exist. In reality student ethnicity is confounded with their achievement status, and minority students should be expected to growth at a higher rate of learning to catch up with their peers.

The study also found that the general category of minority students cannot really capture differences between ethnic groups. The achievement gaps between different minority groups and the reference group (White students) have different statistical characteristics and, thus, may have different policy implications.

One major research question raised in this study is whether methodology matters in terms of analyzing achievement gap. The traditional descriptive approach (using mean score and effect size) and the Hierarchical Linear Modeling were both applied and then compared. The purpose is not to determine which methodology is superior. Rather, the aim is to determine the extent to which the two approaches yield similar or different results. If the traditional descriptive approach gives essentially the same results as the growth modeling, we can argue on behalf of the simpler method. On the other hand, if the two show very different pictures of the achievement gap and its relationship to the characteristics of students, schools, and districts, researchers need to further address these differences.
In our study, the two approaches gave similar results with respect to the general patterns in achievement gaps, but the HLM expanded the analyses to the ethnic difference in growth and had some different details. First, instead of simply comparing mean growth rates, *position of scores* is applied to investigate whether the change of growth rate is related with student initial status. The concept of achievement gap is expanded to the achievement growth gap with the consideration of student initial status. Another major methodological feature of HLM is that *random variability* of multi-level units (scores, students, schools/districts) can be estimated. Instead of including overall variance in effect size analysis, the estimation of random variability of scores in HLM suggests how individual students differed from each other in terms of their achievement and achievement growth as well as how schools or districts differed from each other in terms of the achievement gaps within each school and district.

In contrast to the study by Raudenbush (2004), which compared two approaches to measuring school quality and school improvement, we found more similarities in our results since both approaches in this study were based on the same cohort of students. Researchers and policy makers used to report the trend of achievement gaps based on performance of different student cohorts. One typical example is that the NAEP results are based on same grades but different cohorts of students have been used to report the trend of achievement gaps nationally. The results of this study suggest that models based on a same cohort of students can better control student background characteristics, and hence differences arising from methodologies tend to be smaller. Although the superiority of a true longitudinal study to a repeated cross-sectional study is well-known
when researchers investigate student achievement growth, few studies used the superior approach to analyze achievement gaps. We recommend that in order to obtain a better understanding of achievement gap researchers report both approaches for a comprehensive view.

In addition to the methodological focus, school and district effects were a major research interest of this study. Schools were found to be able to account for a portion of variance in student achievement and achievement growth. Compared with achievement status, schools mattered more as to how fast student scores increased since one third of variance in growth rate was explained by variation among schools. For the Hispanic-White gap, socio-economic status (percentage of FRL students) was a significant predictor of differences in growth rates across schools, while school size (total student number) was a significant predictor of differences in Black-White growth rates across schools. School mobility was not associated with student growth in a school, but it was significantly related with school average achievement. Again, the term “school effect” here has no causal implication, but simply refers to the correlation between a few selected school characteristics and the gaps in achievement and achievement growth across schools. School features can be further investigated and more evidence can be collected to support related policy to narrow achievement gaps or to be used as a confounding factor in the design of a program evaluation.

Compared with schools, districts were found to explain less variance in student achievement and growth rate. However, both the traditional correlation analyses and the HLM models found that the segregation index was a significant predictor of the
achievement gap and growth gap between Hispanic students and White students. In a segregated district where Hispanic students tended to disproportionately cluster in a few schools, the gap between Hispanic students and White students in both achievement status and growth were bigger than the ones in the districts where students of different ethnicities distributed across schools evenly. Orfield (1999) once pointed out that “resegregation” might increase in the first decade of the twenty-first century and it would have an impact on student achievement. The study confirmed that suburban school segregation was significantly associated with student achievement gap and achievement growth gap. More research is needed to be done to confirm this observation.

Returning to the initial research question of whether ethnicity is a meaningful indicator of educational inequality, the research reported here suggests that even if ethnicity is confounded with other indicators, such as initial score and socio-economic status, it remains an important predictor of achievement gaps and achievement growth gaps. Moreover, the distribution of ethnicity at the school level and the district level (as a measure of relative segregation) is related to both student status and growth. Thus, ethnicity should be included in discussions of relevant policies.

In addition to contributing to the debate over the existence and magnitude of achievement gaps and achievement growth gaps between minority students and White students, this study demonstrates that schools and districts matter with respect to reducing the achievement gaps. On the one hand, students in the same schools or districts tend to experience similar patterns in status and growth. On the other hand, in comparison with districts, schools account for a greater proportion of the variability in student growth.
Therefore, school-level policies, as opposed to district-level policies, may be more effective in reducing achievement gaps.

With the aim of eventually eliminating educational inequality, the NCLB act requires schools to disaggregate the student population by race/ethnicity, as well as other characteristics, and to report for each subgroup the percent proficient. If even one subgroup fails to meet its Annual Measurable Objective as determined by the state, the school is labeled as failing to make Adequately Yearly Progress (AYP). Thus, NCLB holds schools accountable not only for the achievement of all its students in the aggregate but also for the achievement of disadvantaged students, such as minority students, students with limited English proficiency, or students with disabilities.

However, schools with greater diversity are more likely to fail to make AYP because they have more subgroups for which to be held accountable. Equally important, disadvantaged students are more likely to have an initial status well below the proficient level and so have greater difficulty achieving grade level proficiency. Consequently, NCLB is often criticized for unfairly judging schools with heterogeneous student populations. To address this criticism different growth models have been introduced by some states so that schools are also judged by how much progress their students make. The present study offers an alternative by exploring how schools might be judged by how much progress they make in reducing achievement gaps. Growth rates of two groups (minority students and White students, in our study) can be measured, compared, and evaluated. If an indicator related to achievement growth gaps is incorporated in a school
accountability system, then goal of reducing educational inequality between groups can be explicitly represented in the system.

Although the study primarily focuses on average achievement growth rates and average growth gaps, the distributions of achievement growth rates and growth gaps can always be obtained from the models employed in this study. The main purpose would not be to rank students but, rather, to inform teachers and schools about how their students’ achievement growth patterns stand in relation to other students, as well as how large the growth gaps are in their organizations compared with other organizations. In principle, teachers could set reasonable targets for their students based on their initial achievement and other characteristics. The average growth rate and the average growth gap for each teacher or each school may be used as a basis for their evaluation in comparison to other teachers or schools. Such comparisons should be made holding relevant teacher or school characteristics fixed. However, the HLM results are statistical descriptions and their use in causal attribution should be done carefully and in conjunction with other relevant evidence.

The Limitations of the Study and the Potential Directions for Future Studies

In view of the fact that the analyses reported here are based on a state assessment system, one critical aspect of the study is test quality. First, there is a question of whether the test is capable of measuring students’ knowledge and skills accurately and precisely for the three grades of interest. A second question is whether the tests, across years and grades, have been appropriately equated and linked. This is crucial as the state and the testing contractor claim that student growth can be measured and our analyses rely on
that claim. We have discussed that the choice of different testing models, in particular Item Response Theory (IRT) and Classical Test Theory (CTT), may have an impact on the variance of the scores. This study thus has to rely on the assumption that the state assessment is well designed to assess student knowledge and skills with adequate validity. Evaluating this assumption was beyond the scope of this study. However, for, by making this assumption the study focused on issues of ethnic differences in test score trajectories. It would be worthwhile to explore how test validity issues including the application of different testing models could influence the analyses of the achievement gap using the same methodologies employed in this study.

Another limitation is that although three years of scores are sufficient for a HLM repeated measure analysis, calculation of growth rates could be sensitive to measurement uncertainty and, obviously, the assumption of linearity. More data points can be sought in the future study to have a more stable prediction of student growth patterns. When more years of achievement are to be tracked, student mobility will become a more salient issue. When students are nested in different schools and districts over years, the cross-classified HLM is a possible methodological approach.

The small number of Black students in the study is of concern. Observations and analyses of the achievement gap and the achievement growth gap between Black students and White students should be interpreted cautiously since they are likely sensitive to small sample fluctuations. Last, although we adopt the word “school effect” in this study to be consistent with the current literature, we are very much aware that the school effect is more of an estimate of the residuals at different levels; it depends on which variables
are put in the model and thus may contain large standard errors and bias. The word effect here has no causal implication.

In sum, we suggest that generalizations of the findings and the conclusions of this study may not be warranted as they are based on one state assessment system and a selected set of suburban schools and districts. We are interested in further investigating the trend of the achievement gaps as well as their relations with school and district characteristics at the national level by replicating the two approaches used in this study.
References:


Appendix A: Analysis for Math Grade Progression

One cohort of students is tracked in this dissertation through the three academic years. Their grade progressions ideally should be reported as grade 6 in 2003, grade 7 in 2004, and grade 8 in 2005, which is coded as 60708. Students with bizarre sequence of grades might be excluded from analysis. For example, a student might be reported in grade 6 in 2003, grade 7 in 2004, and grade 10 in 2005 (coded as 60710). It’s necessary to examine all possible grade progressions and sort out those with abnormal progressions. Below is a table of all the grade progressions observed for the math test from grade 6 to grade 8—the first one digit refers to the grade in 03, the second two digits refer to the grade in 04, and the third two digits refer to the grade in 05. The rule to be used in determining which grade progressions get analyzed and which don’t is that we will only utilize a reasonable sequence of grades. Moreover, we will allow scores in a given grade for a given year to be used as proxies within that grade and any other year. We will detail our decision on whether to delete or keep each grade progression in the table below (see Table 1).

Two facts should be born in mind before examining grade progressions: First, some students only took the tests through two years instead of three years. Even without the complete three-year records, they will still be kept in the dataset since two scale scores construct a projection. Therefore, the progressions, such as 00708 and 60700, will be remained in the dataset. Second, some odd progressions could be due to the possibility that two different students were matched over years. The records with non-traditional grade progressions must either be deleted from the data or “corrected” to fit one of our
targeted progressions (50607). For example, grade progression 60710 cannot be analyzed since the 10th grade score would have to be considered as a 8th grade score. A few other progressions are more complicated. For example, a student with a 70708 grade progression in math can be either due to a random error or because this student stayed in the same grade for two years. Although it is reasonable for us to believe that the test the student took yields a scale score equivalent to what they would have taken had they done the other test, considering such cases are few (less than 0.1 percent), we decided to delete them (see Table A1.)

Table A1

<table>
<thead>
<tr>
<th>Progression</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Decision Making</th>
</tr>
</thead>
<tbody>
<tr>
<td>00708</td>
<td>1357</td>
<td>4.2</td>
<td>4.2</td>
<td>K (Keep)</td>
</tr>
<tr>
<td>60005</td>
<td>1</td>
<td>.0</td>
<td>.0</td>
<td>D (Delete)</td>
</tr>
<tr>
<td>60006</td>
<td>2</td>
<td>.0</td>
<td>.0</td>
<td>D</td>
</tr>
<tr>
<td>60007</td>
<td>33</td>
<td>.1</td>
<td>.1</td>
<td>D</td>
</tr>
<tr>
<td>60008</td>
<td>598</td>
<td>1.8</td>
<td>1.8</td>
<td>K</td>
</tr>
<tr>
<td>60009</td>
<td>20</td>
<td>.1</td>
<td>.1</td>
<td>D</td>
</tr>
<tr>
<td>60508</td>
<td>1</td>
<td>.0</td>
<td>.0</td>
<td>D</td>
</tr>
<tr>
<td>60600</td>
<td>22</td>
<td>.1</td>
<td>.1</td>
<td>D</td>
</tr>
<tr>
<td>60607</td>
<td>161</td>
<td>.5</td>
<td>.5</td>
<td>D</td>
</tr>
<tr>
<td>60608</td>
<td>6</td>
<td>.0</td>
<td>.0</td>
<td>D</td>
</tr>
<tr>
<td>60609</td>
<td>3</td>
<td>.0</td>
<td>.0</td>
<td>D</td>
</tr>
<tr>
<td>60700</td>
<td>1275</td>
<td>3.9</td>
<td>3.9</td>
<td>K</td>
</tr>
<tr>
<td>60705</td>
<td>1</td>
<td>.0</td>
<td>.0</td>
<td>D</td>
</tr>
<tr>
<td>60707</td>
<td>188</td>
<td>.6</td>
<td>.6</td>
<td>D</td>
</tr>
<tr>
<td>60708</td>
<td>28741</td>
<td>88.0</td>
<td>88.0</td>
<td>K</td>
</tr>
<tr>
<td>60709</td>
<td>29</td>
<td>.1</td>
<td>.1</td>
<td>D</td>
</tr>
<tr>
<td>60710</td>
<td>2</td>
<td>.0</td>
<td>.0</td>
<td>D</td>
</tr>
<tr>
<td>60800</td>
<td>4</td>
<td>.0</td>
<td>.0</td>
<td>D</td>
</tr>
<tr>
<td>60808</td>
<td>1</td>
<td>.0</td>
<td>.0</td>
<td>D</td>
</tr>
<tr>
<td>60809</td>
<td>22</td>
<td>.1</td>
<td>.1</td>
<td>D</td>
</tr>
<tr>
<td>60910</td>
<td>2</td>
<td>.0</td>
<td>.0</td>
<td>D</td>
</tr>
<tr>
<td>70708</td>
<td>183</td>
<td>.6</td>
<td>.6</td>
<td>D</td>
</tr>
<tr>
<td>Total</td>
<td>32652</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>
In sum, although there are only a few progressions that are analyzed, most of the records (98%) are actually kept (see Table A2).

Table A2
Math Grade Progression for 2003, 2004, 2005

<table>
<thead>
<tr>
<th>Progression</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>00708</td>
<td>1357</td>
<td>4.2</td>
<td>4.2</td>
<td>4.2</td>
</tr>
<tr>
<td>60008</td>
<td>598</td>
<td>1.9</td>
<td>1.9</td>
<td>6.1</td>
</tr>
<tr>
<td>60700</td>
<td>1275</td>
<td>4.0</td>
<td>4.0</td>
<td>10.1</td>
</tr>
<tr>
<td>60708</td>
<td>28741</td>
<td>89.9</td>
<td>89.9</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>31971</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>