2013

Student Achievement in Middle Grades: Gauging the Effect of Teacher Training on Student Learning

Natalya Gnedko
Loyola University Chicago, nsgnedko@gmail.com

Recommended Citation
Gnedko, Natalya, "Student Achievement in Middle Grades: Gauging the Effect of Teacher Training on Student Learning" (2013). Dissertations. Paper 666.
http://ecommons.luc.edu/luc_diss/666

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Copyright © 2013 Natalya Gnedko
LOYOLA UNIVERSITY CHICAGO

STUDENT ACHIEVEMENT IN MIDDLE GRADES:
GAUGING THE EFFECT OF TEACHER TRAINING
ON STUDENT LEARNING

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

PROGRAM IN RESEARCH METHODOLOGY

BY

NATALYA S. GNEDKO

CHICAGO, IL

AUGUST 2013
ACKNOWLEDGEMENTS

A dissertation is a journey that cannot be completed without the help and support of many people. I would like to thank those who were with me along the way and without whom I would have been lost. I would like to thank Dr. Meng-Jia Wu for being a positive force and for believing in me more than I sometimes believed in myself. I would like to thank Dr. Terri Pigott for the never ending support even at the time of personal and professional changes. I would like to thank Dr. Kelci Price for being my voice of reason and encouragement that I heard loud and clear despite the many miles between us. I am very grateful to my husband, Brian Berry, for his support of this lengthy endeavor and for giving me the much needed time. I also very much appreciate my daughter, Julia Berry, for often allowing me to retreat in my corner and keep on typing, even though she wondered about the wisdom of engaging in something that clearly looked less interesting than playing with her. Last but not least, I would like to thank my parents for their lifelong support of my quest for knowledge, even at the time when they quietly questioned the way in which I chose to spend nearly all of my thirties.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ...................................................................................................................... iii

LIST OF TABLES ..................................................................................................................................... vi

LIST OF FIGURES .................................................................................................................................. vii

LIST OF ABBREVIATIONS ................................................................................................................ viii

ABSTRACT ................................................................................................................................................ ix

CHAPTER ONE: TEACHER QUALITY: A FACTOR IN STUDENT ACHIEVEMENT? ...................... 1
  Current Study ........................................................................................................................................ 8

CHAPTER TWO: WHAT WE KNOW AND WHAT WE NEED TO LEARN ABOUT
  TEACHER TRAINING AND STUDENT ACHIEVEMENT ......................................................... 11
  Teaching Certificates as a Measure of Teacher Quality ................................................................. 11
    Teaching Certificates by Type ....................................................................................................... 14
    Teaching Certificates by Content Area ...................................................................................... 19
  College Background as a Measure of Teacher Quality ................................................................. 24
  Research Limitations and Contributions of Current Study ......................................................... 30

CHAPTER THREE: METHOD ............................................................................................................ 34
  Sample ................................................................................................................................................ 34
  Data and Instrumentation ................................................................................................................. 35
    Data .............................................................................................................................................. 35
    Instrumentation ............................................................................................................................. 37
  Procedure .......................................................................................................................................... 39
  Analysis ............................................................................................................................................ 40
    Exploratory Analysis .................................................................................................................. 40
    Multilevel Growth Analysis ........................................................................................................ 41

CHAPTER FOUR: RESULTS ............................................................................................................... 47
  Sample Characteristics .................................................................................................................. 47
  Exploratory Analysis of Student Achievement and Teacher Training ........................................ 50
    Student Achievement ................................................................................................................. 50
    Teacher Training .......................................................................................................................... 51
  Relationship Between Achievement, Training, and Other Study Variables ......................... 54
  Multilevel Growth Analysis ............................................................................................................ 57
    Math ............................................................................................................................................. 57
    Reading ....................................................................................................................................... 64

CHAPTER FIVE: DISCUSSION ........................................................................................................... 70
LIST OF TABLES

Table 1: Records included in the study............................................................................. 35
Table 2: Sample characteristics across middle grades and schools ......................... 48
Table 3: Middle grades test scores over time................................................................. 50
Table 4: Results of one-way repeated measures ANOVA for test scores ................. 51
Table 5: Across-school distribution of endorsed teachers teaching in-field.............. 52
Table 6: Yearly counts of in-field teachers with middle grades endorsements ........ 53
Table 7: Results of one-way repeated-measures ANOVA for endorsements ............ 54
Table 8: Results of Pearson product-moment correlation analysis
     between mean test scores and endorsements............................................................. 55
Table 9: Results of Pearson product-moment correlation analysis
     between average test scores and control variables ................................................. 56
Table 10: Results of multilevel growth analysis in math
     (standard errors in parentheses)........................................................................... 59
Table 11: Results of multilevel growth analysis in reading
     (standard errors in parentheses)........................................................................... 66
Table 12: Yearly counts of teachers with middle grades endorsements
     teaching in-field and out-of-field............................................................................ 76
LIST OF FIGURES

Figure 1: Middle grades test scores over time ............................................................... 50
Figure 2: Yearly counts of in-field teachers with middle grades endorsements.......... 53
Figure 3: Math achievement over time at schools with and without math-endorsed teachers ............................................................................................. 60
Figure 4: Percent of math- and reading-endorsed teachers assigned to teach in-field .................................................................................................. 77
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPS</td>
<td>Chicago Public Schools</td>
</tr>
<tr>
<td>ITBS</td>
<td>Iowa Test of Basic Skills</td>
</tr>
<tr>
<td>LSAY</td>
<td>Longitudinal Study of American Youth</td>
</tr>
<tr>
<td>MET</td>
<td>Measures of Effective Teachers Project</td>
</tr>
<tr>
<td>NAEP</td>
<td>National Assessment of Educational Progress</td>
</tr>
<tr>
<td>NCLB</td>
<td>No Child Left Behind</td>
</tr>
<tr>
<td>NELS</td>
<td>National Educational Longitudinal Study of 1988</td>
</tr>
<tr>
<td>OLS</td>
<td>Ordinary Least Squares</td>
</tr>
<tr>
<td>SAT-9</td>
<td>Stanford Achievement Test, Ninth Edition</td>
</tr>
<tr>
<td>TAP</td>
<td>The System for Teacher and Student Advancement</td>
</tr>
<tr>
<td>TFA</td>
<td>Teach for America</td>
</tr>
</tbody>
</table>
ABSTRACT

In the United States, teacher certification has been a baseline measure of teacher quality and gateway to the teaching profession for many decades. Research suggests that teacher certification is beneficial to student achievement, with findings particularly promising when content area of teacher certification is taken into account. Positive findings are largely limited to high school. However, in recent years middle grades have been increasingly viewed as a starting point for success in high school and college, which led many school districts to require advanced training for middle grades teachers. To examine the effectiveness of advanced teacher training in middle grades, the study analyzed the nature of longitudinal relationship between student outcomes and teacher training in middle grades math and reading at one of the largest public school districts in the Midwest. Using a multilevel model of growth at the school-level of analysis, the study found a positive effect of advanced teacher training in math on student achievement in math. The study did not find evidence of the positive effect of teacher training in reading. Examination of the local policy context suggested that changes in the district could have impacted the study's ability to estimate teacher effect in reading.
CHAPTER ONE

TEACHER QUALITY: A FACTOR IN STUDENT ACHIEVEMENT?

In the United States, the quality of public education remains a topic of heated debate. In calling the state of public education a “catastrophe”, the documentary Waiting for “Superman” (Weyermann & Guggenheim, 2010) captured the sentiment of many parents and students involved in a public school system, particularly in the large urban districts that serve a high percent of minority and low-income students. Federal and local governments, education authorities, research institutions, parents and teachers are among the groups involved in the discussion about what can be done to improve student achievement in public schools. This discussion often shifts to the quality of teachers, with general public and policy makers in agreement that good teachers are essential to student success (e.g., Green, 2010; Harris, 2010; Hess, 2006; Kristof, 2006).

If teachers are important to student learning, then it is reasonable to expect that improvement in teacher quality should lead to improvement in student achievement. This logic fueled a series of national and local efforts aimed at improving teacher workforce. For example, the No Child Left Behind (NCLB) legislature of 2001 has teacher quality as one of its provisions. The legislature requires that all teachers become highly qualified: have bachelor degrees, full state certification or licensure, and prove that they are experts in the subject they teach.
The Obama administration stated that it would increase funding for NCLB (Chait, 2009). It is likely that at least a portion of funds will be available for increasing the ranks of highly qualified teachers. In a more recent effort, the federal government allocated $4.35 billion for the Race to the Top Fund as part of the American Recovery and Reinvestment Act. Race to the Top is a competitive program designed to reward states that encourage innovation and reform in education (U.S. Department of Education, 2009). One area of the program’s focus is teaching workforce: states must recruit, develop, reward, and retain effective teachers and principals. The program guidelines require the effectiveness of teachers and principals to be evaluated based on student performance (U.S. Department of Education, 2009). The Investing in Innovation Fund, also established under the American Recovery and Reinvestment Act, provides competitive grants to applicants who implement innovative practices that improve student achievement (U.S. Department of Education, 2009). Under this initiative funds have been awarded for the implementation of programs aiming at improving teacher quality. For example, Boston Plan for Excellence in the Public Schools Foundation was among the top rated 2010 proposals with the plan to develop and retain highly effective teachers in the areas of greatest need in Boston Public Schools (U.S. Department of Education, 2010).

The federal government is not the only entity taking concrete steps to improve teacher quality. Private organizations have long devoted funds and efforts to the development of teaching workforce. For example, in 1999 the Milken Family
The Bill and Melinda Gates Foundation recently announced a plan to provide all students with education that would prepare them for college. Recognizing that teachers are essential in achieving this goal, the foundation also announced the Measures of Effective Teachers Project (MET). The goal of the project is to understand what makes teachers effective. MET researchers plan to develop fair and reliable measures of teacher effectiveness that are not solely based on student test scores. Instead, they seek to take into account a full range of teachers’ work and the context in which they teach. The project began in the ’09-’10 school year at a number of school districts throughout the nation (Bill and Melinda Gates Foundation, 2011).

Local groups and organizations echo the efforts of large foundations and the federal government. Although their initiatives are smaller in scale, combined they reach many students and teachers in school districts nationwide. For example, in 1998 the Fund for Educational Excellence began offering professional development and instructional support to Baltimore literacy teachers. The goal of this initiative, called Achievement First, is to improve students’ achievement in literacy by
investing in teacher preparation (Fund for Educational Excellence, 2011). From 2006 to 2010, thirty-four public schools in Chicago participated in the Cluster 4 Middle Grades Project. The project, initially funded by the Chicago Community Trust, was implemented in order to improve math and science education in middle grades. Improving quality of math and science teachers was one of the ways the project aimed at achieving this goal. As part of the program participation, teachers received professional support in math and science as well as stipends for taking math and science courses at the local universities (Chicago Public Schools, 2011; Middle Grades Forum, 2009).

The above examples illustrate the variety and the range of resources devoted to the development of teaching workforce. Is the investment in teachers paying off? Some initiatives that allocate resources to teacher quality do not directly measure the impact of investment on student achievement. For example, under NCLB teachers have to become highly qualified, but the legislature does not require teacher qualifications to be measured against student outcomes. Instead, NCLB requires that schools as a whole make adequate yearly progress (U.S. Department of Education, 2004). More recent programs, such as Race to the Top and Investing in Innovation, require teacher effectiveness to be evaluated based on student achievement. However, these programs are new and haven’t yet accumulated research evidence. Findings from smaller, local initiatives are often not published, leaving their effectiveness largely unknown. Meanwhile, millions of teachers go to the classrooms every day. How do we know that all teachers, those who do and do
not have an opportunity to participate in professional development programs, enter
their classrooms prepared to teach?

For decades, teacher certification has been a gateway to the teaching
profession. Although often a contentious issue, teacher certification remains the
basic measure of teacher quality with the core requirements similar across the
states (Darling-Hammond, 1999). Educational researchers have long been
interested in the effectiveness of teacher certification. Some examined teacher
certification by comparing student outcomes between fully certified teachers
(teachers who completed all requirements for a standard certificate) and under-
certified teachers (teachers with temporary or emergency certificates who did not
complete all requirements for a standard certificate). The findings in this line of
research are mixed. While some studies present evidence of a positive effect of full
teacher certification on student learning (e.g., Clotfelter, Ladd, & Vigdor, 2007;
Darling-Hammond, 1999; Darling-Hammond, Holtzman, Gatlin, & Vasquez Heilig,
2005; Greenberg, Rhodes, Ye & Stancavage, 2004; Laczko-Kerr & Berliner, 2002),
other studies find no significant difference in student outcomes between teachers
who are fully certified and teachers who are not (e.g., Decker, Mayer, & Glazerman,
2004; Kane, Rockoff, & Staiger, 2008).

A more targeted approach to studying the effectiveness of teacher
certification is to take into account content area of certificates (e.g., whether a
certificate is in math, science, reading, or another subject). Researchers who used
this approach typically compared student outcomes between teachers certified to
teach their subject and those teaching outside of their area of expertise. The majority of studies have focused on math and science. The results suggest that teachers’ expertise in the subject that they teach is beneficial to student learning. For example, in a series of studies Goldhaber and Brewer found a positive relationship between teacher certification in math and science and student achievement in math and science in grades 10 and 12 (Goldhaber & Brewer, 1996; Goldhaber & Brewer 1997; Goldhaber & Brewer, 2000). Hawk, Coble, and Swanson (1986) and Darling-Hammond (1999) also found evidence of the positive relationship between teachers’ content area certification and student outcomes in math and to some extent in reading. A group of researchers studying the effectiveness of content area training used teachers’ college background instead of the certificates. Results of their studies also suggest that in math and science teacher training in the subject that they teach is beneficial to student learning (e.g., Aaronson, Barrow, & Sander, 2007; Chaney, 1995; Greenberg et al., 2004; Monk, 1994; Monk & King, 1994; Rowan, Chiang, & Miller, 1997; Wenglinsky, 2000).

Generally, studies that examined the effectiveness of teacher training in the area that they teach found a positive relationship between training and student outcomes. These studies, however, share a few shortcomings. For example, researchers devoted little attention to subjects other than math and science that could be important to academic advancement, such as reading. In addition, researchers mostly focused on high school. The emphasis on high school is likely driven by the expectation that high school teachers should have content area
expertise in order to teach advanced classes to college-bound students (USAA Educational Foundation, 2010; U.S. Department of Education, 2009). However, recently the U. S. Department of Education has elevated the importance of middle grades learning, describing middle grades as a starting point for college readiness and advising that students begin taking advanced courses, such as Algebra, in grades 7 and 8 (U.S. Department of Education, 2010).

Many school districts echo the U. S. Department of Education’s recommendation by requiring that teachers receive specialized training if they want to teach core subjects in middle grades (usually math, science, social studies, and language courses). For example, in 2008 Chicago Public Schools (CPS) Board of Education adopted the Middle Grades Specialization Policy that requires middle grades teachers who want to teach core subjects to complete training in content area and pedagogy relevant to middle grades (Chicago Public Schools, 2008; Chicago Public Schools, 2011). The idea of setting higher standards for middle grades teachers is resonating throughout the state of Illinois, where CPS is located, with the state government considering statewide changes similar to the CPS policy (Harris, 2010). Requirements similar to CPS appear in other school district that also expect that middle grades teachers receive advanced training in content area and education (e.g., District of Columbia Office of the State Superintendent of Education, 2009; New York State Education Department, 2011; State of California Commission on Teacher Credentialing, 2011).
While the demand for content area experts in middle grades has been increasing, few researchers examined the effectiveness of teacher preparation in middle grades – a fact that has been previously acknowledged. For example, after a review of 16,000 research publications and policy reports, the National Mathematics Advisory Panel concluded that the evidence of relationship between teacher training and student outcomes in middle grades had been difficult to find and that the existing studies had been methodologically flawed (U. S. Department of Education, 2008). After conducting a large-scale middle grades study in 303 schools in California, Williams, Kirst, Haertel, et al. (2010) also pointed out the difficulty of finding research that focused on middle grades. In preparation of this review, I was able to locate a small group of studies that examined the effectiveness of middle grades teachers. These studies offer some support to the idea that middle grades students benefit academically from teachers with advanced training (Chaney, 1995; Darling-Hammond, 1999; Greenberg et. al., 2004; Hawk et. al., 1986; Wenglinsky, 2000). However, methodological differences between the studies are substantial and include analytic techniques, approach to data aggregation, and ways in which teacher preparation was measured.

**Current Study**

The goal of the current study is to expand what is known about the effectiveness of advanced teacher training in middle grades. The study examines the nature of the relationship between teacher training and student outcomes in middle grades at one of the largest school districts in the Midwest. The study examines
school-level outcomes in math and reading in grades six through eight during the
school years 2007 through 2011.\textsuperscript{1} The selection of subjects in the current study
complements earlier research that has traditionally focused on math and expands
this research by including reading.

The study examines school-level outcomes using a multi-level model of
growth. Using a multilevel model of growth allows for a longitudinal analysis of the
relationship between teacher training and student achievement. This approach to
analysis expands on the methods used in earlier research on teacher effectiveness,
which traditionally relied on multiple regression of cross-sectional data (e.g., Boyd
et al., 2007; Chaney, 1995; Darling-Hammond, 1999; Greenberg et al., 2004),
sometimes controlling for prior achievement to examine yearly gains (e.g., Clotfelter
et al., 2007; Darling-Hammond et al., 2005; Decker et al., 2004; Kane et al., 2008;
Monk, 1994; Monk & King, 1994). The study addresses the following research
questions:

\textit{Question 1:} What is the pattern of change in teacher training and student outcomes
in middle grades math and reading at the district under study?

\textit{Question 2:} What is the nature of longitudinal relationship between school-level
outcomes in middle grades’ math and reading and teacher training in math and
reading in the district under study?

\textsuperscript{1} Throughout this paper, school years are referenced based on the year in the spring semester. For
example, the school year ’10-’11 is referenced as 2011.
In the educational climate of increasing demand for advanced teacher training in middle grades, results of the current study may be of interest to policy makers and school officials across the nation, including school districts that have already introduced higher standards for middle grades teachers and those that are considering similar changes. The results may also be of interest to the existing and prospective middle grades teachers who could be impacted by the changes in teaching requirements. Lastly, current and future middle grade students and their families may be interested in the study’s results, as they could be directly affected by the shifting educational policy.

In the following chapters, I take a closer look at what is currently known about the effectiveness of teacher training and addresses mechanics of the current study. Chapter 2 offers a detailed summary of previous research on teacher effectiveness. Chapter 3 describes the study’s design and methodology. Results are summarized in Chapter 4, and Chapter 5 offers a discussion of findings and outlines the study’s limitations and directions for future research.
CHAPTER TWO

WHAT WE KNOW AND WHAT WE NEED TO LEARN ABOUT TEACHER TRAINING AND STUDENT ACHIEVEMENT

The goal of this chapter is to describe the earlier research that examined the relationship between teacher training and student learning. The focus of the review is primarily on the studies that measured teacher training using teacher certification. The decision was made to focus on certification, because it remains a basic measure of teacher quality in public education across all states. The review also includes a group of studies that measured training using college preparation in a subject matter and in education. These studies were included, because preparation in a subject matter and in education are the two core elements of teacher certification, making results of these studies comparable to the ones that used certificates. In conclusion, the chapter provides a summary of research findings and limitations and describes how the proposed study aims to amend the gap in what we currently know about the effectiveness of teacher training in middle grades.

Teaching Certificates as a Measure of Teacher Quality

Across all states, teaching certificates are used to ensure that people entering the teaching profession meet the basic standard of quality. To be fully certified, teaching candidates must satisfy all requirements set forth by the board of
education in a state where they want to teach. While there are variations among the states in the specifics of certification, the core standards are similar. For example, teachers in all states must hold a bachelor degree or a graduate degree. Most states ask for a major or a minor in the field to be taught plus an average of thirty credit hours in education. In many states teaching candidates have to take pre-certification tests (e.g., one or more tests of basic skills, test of subject matter knowledge, test of teaching knowledge). Student teaching is also common. After satisfying all requirements, new teachers usually receive an initial certificate that can be upgraded to a standard certificate after a probationary teaching period (e.g., four years in Illinois). Standard certificates usually have to be renewed every five years after completion of a professional development course (for a summary see Darling-Hammond, 1999; for links to the individual states’ guidelines see the University of Kentucky College of Education, 2011).

The requirements for becoming a fully certified public school teacher seem rigorous. However, many public schools lag behind in academics. Other reasons, such as family environment and school climate, could affect student performance regardless of teacher preparation. Yet could it also be that teaching certification misses the mark on ensuring quality? A group of researchers and people involved in the education debate think exactly that. They share the belief that certification requirements, particularly the need for education courses, discourage talented individuals from becoming teachers. They argue that successful professionals and college graduates could teach just as well or better than fully certified teachers (e.g.,
Hess, 2006; Kristof, 2006; Leigh & Mead, 2005). In reality, in most states full certification is one of a number of pathways for becoming a teacher. Often teachers hold alternative certificates. Alternative certificates are usually issued to teaching candidates who did not complete a traditional certification program, but who can demonstrate college or professional background that would have sufficiently prepared them for teaching. Preparation in education is typically not required to receive an alternative certificate (Fenstermacher, 1990). For example, Teach for America (TFA) is a well-known alternative certification program. TFA participants, who are usually advanced college graduates, are recruited for two-year teaching commitments in low-income communities. TFA members do not complete a teaching program before becoming teachers and do not usually hold a standard teaching certificate when they begin teaching. Instead, they participate in an intensive five-week training program prior to beginning a teaching assignment (Teach for America, 2011). In addition to alternative certificates, many states issue emergency or temporary teaching certificates. Unlike alternative certificates, emergency and temporary certificates usually require little or no preparation prior to the beginning of a teaching assignment and are usually valid from six months to two years. School districts often rely on these certificates at the time of teacher shortages (Fenstermacher, 1990). Although little formal preparation is officially required for an emergency or a temporary certificate, sometimes these certificates are issued to experienced teachers who are in the process of completing requirements for a standard certificate (Darling-Hammond, Berry, Thoreson, 2001).
Since different certificate types are used to put teachers in the classrooms, the question about the value of full certification may be addressed by comparing the effectiveness of certificate types, which was done in the studies summarized in the following section.

*Teaching Certificates by Type*

The most common types of teaching certificates are the standard certificates (often referred to as the full certificates), alternative certificates, and the emergency (temporary) certificates. Standard certification is associated with the most rigorous list of requirements, including preparation in the subject matter, education, and often testing by the state. Does the highest standard in teacher preparation lead to the highest student achievement? A number of studies suggest that full certification is beneficial. For example, Darling-Hammond (1999) found evidence in support of full certification in the state-level analysis of the National Assessment of Educational Progress (NAEP) data from 44 states. The data included scores in math and reading at the 4th grade level. At the 8th grade level, only math scores were available. Each grade included two years of data. Darling-Hammond, similarly to other researchers who examined NAEP data, used the Schools and Staffing Survey of ’93-’94 to access teacher characteristics and link them with the NAEP data. The results showed that the percent of new teachers who were not fully certified correlated negatively with student achievement in all subjects, all grades, and all years. Controlling for student and school characteristics, the partial correlation coefficients ranged from -.40 to -.63. The relationship between the percent of teachers with standard certificates and
student achievement was positive, with the coefficients ranging from .20 to .57. Darling-Hammond concluded that students learn better when their instructors are fully certified. Greenberg et. al. (2004) analyzed the 8th grade NAEP data but at the student-level. The researchers classified teachers as certified if they had advanced, standard, or probationary\(^1\) teaching certificates. All other teachers, including teachers with the emergency, temporary, or provisional certificates were classified as not certified. Controlling for student ability, school characteristics, and family background, the researchers found that math achievement of the 8th grade students was approximately 8-points higher if they were taught by certified teachers. Neither study controlled for the prior student achievement, as this information was not included in the NAEP files. Using data from California public schools, Fetler (1999) found that students’ math scores on the Stanford Achievement Test, Ninth Edition (SAT-9) were negatively associated with the percent of teachers who taught math under the emergency certificates. Standard estimates varied from -3.1 to -4.1. The analysis was based on 795 schools and included grades 9 through 11. Fetler also did not control for the prior achievement.

To examine the differences between fully certified and under-certified teachers, Laczko-Kerr and Berliner (2002) matched the two groups of teachers based on the similarities of their schools and districts. The matching procedure yielded 109 pairs. All teachers in the study were from five school districts in  

---

\(^1\) Probationary certificates are issued to new teachers who met all certification requirements, but have not yet completed a probationary teaching period.
Arizona. The analysis of variance showed that in 1998 students in grades 3 through 8 who were taught by teachers with standard certificates had significantly higher reading and language scores on SAT-9 than students who were taught by under-certified teachers. In 1999, the results also included math. A separate analysis of student outcomes between fully certified teachers and TFA recruits showed that students of fully certified teachers performed significantly better than students of TFA recruits in all subjects in 1999. The difference between the two groups was not significant in 1998, however the same pattern of results was present.

Controlling for the prior student achievement allows for the examination of teacher effect on gains in student performance. A group of studies that controlled for the prior achievement found evidence of a positive relationship between full teacher certification and gains in student learning. However, the relationships found in these studies were small. For example, Darling-Hammond et al. (2005) conducted a student-level analysis of the 4th and 5th graders' test scores in math and reading in Texas. The researchers analyzed results from ’95-’96 to ’01-’02 on three different tests: the Texas Assessment of Academic Skills, SAT-9, and the Aprenda test. The results showed that relative to the teachers with standard certificates, uncertified teachers and teachers with alternative or emergency certificates had a negative effect on student gains on nearly all math and reading tests. While the majority of the reported unstandardized coefficients were significant, their size was relatively

---

2 Aprenda is used to measure academic achievement of Spanish-speaking students in grades K-12. It is modeled after the Stanford Achievement Test series.
small, ranging from -2.32 to -.45. The researchers ran a separate analysis to compare student outcomes between fully certified teachers and TFA recruits. The results showed that uncertified TFA recruits had a significant negative effect on gains in student performance, unstandardized coefficients ranging from -6.12 to -.97. The authors concluded that under-certified teachers had a negative impact on student learning. Clotfelter et al. (2007) examined math and reading test scores in North Carolina from 1995 to 2005. The researchers found that relative to teachers with standard certificate, teachers with other types of certificates (e.g., provisional, emergency, temporary certificates) had a significant negative effect on student gains in grades 4 and 5 in both math and reading. The effect was small at -.06 standard deviations in math and -.02 standard deviations in reading.

While the results of the studies reviewed so far suggest that full teacher certification is beneficial to student learning, this is not always the case. Decker et al. (2004) examined the effectiveness of TFA recruits and other teachers (with standard certificates, alternative certificates, or no certificates) from 17 schools across five large school districts. Approximately 2,000 students in grades 1 through 5 participated in the study. The students were randomly assigned to classrooms taught by TFA non-TFA teachers. Using the Iowa Test of Basic Skills (ITBS), the researchers assessed students’ knowledge of math and reading at the beginning and at the end of school year. The results showed that gains in math outcomes of students taught by TFA recruits were approximately .15 standard deviations higher than gains of students taught by other teachers. When they were compared to
teachers with standard certificates, the impact of TFA recruits was slightly lower. However, the researchers stated that the subgroup sample sizes were too small and the standard errors were too high to draw inferences. The results in reading showed no significant differences between the TFA recruits and other teachers.

The study by Kane et al. (2008) is another example that casts doubt on the benefit of full certification. Using six years of the NYC math and reading test data for students in grades 4 through 8, the researchers compared student outcomes between teachers with standard certificates, teaching fellows, TFA recruits, teachers from international programs, and uncertified teachers. They analyzed data at student-level, with teacher and student records linked. Predictive models were based on the records of approximately 10,000 students. The results showed that on average, full certification had little impact on gains in student achievement in both reading and math: the difference between teachers with standard certificates and teachers with other certificates was .05 standard deviation or less. A separate analysis of data that did not control for the prior achievement showed that students taught by teachers with standard certificates outperformed students taught by all other teachers. The largest difference was between certified teachers and international recruits in math at .48 standard deviations.

While there is evidence in favor of full teacher certification, some well designed and controlled studies present mixed findings (Decker et al., 2004; Kane et al., 2008). Grouping teachers based solely on the certificate type may be one reason for unclear results, as this way of grouping does not account for the content area of
teacher training. Including content area in the analysis may not be possible in primary grades, where teachers are usually certified to teach all subjects in a self-contained classroom. However, teachers in secondary grades\(^3\) are typically certified in one or more areas, such as math, history, or English, something that is frequently missed in the design of studies. For example, Laczko-Kerr and Berliner (2002) compared student outcomes between all teachers with standard certificates and all teachers with other certificates. The group of teachers with standard certificates included teachers who majored in education, liberal arts, special education, physical education, and other fields. Student outcomes included math, reading and language: all subjects not related to teachers’ training. While it is informative to know that there may be an overall positive effect of full teacher certification, the relationship between certificates and student achievement could be examined more accurately by taking content area into account, which was done in the studies described in the following section.

*Teaching Certificates by Content Area*

By and large, the results of the studies that examine the relationship between content area certification and student outcomes suggest that students benefit from teachers who teach within their area of certification. In one of the earlier studies, Hawk et al. (1986) analyzed math achievement of 826 CPS students in grades 6 through 12. Of the thirty-six teachers who participated in the study, half of the

\(^3\) Although there is some variation between school districts, grades 1 – 5 are usually considered primary grades, and grades 6 through 12 are usually considered secondary grades. Secondary grades include middle school (grades 6-8) and high school (grades 9-12).
teachers were certified in math and half were certified in other areas. Teachers teaching within their content area were paired with teachers teaching outside their content area at the same school. Paired teachers were teaching the same course to students of similar ability. Using Stanford Achievement Test (SAT) scores to measure student ability in general math and algebra, the researchers conducted the pre- and post-tests five months apart. The analysis of variance showed no significant differences in student achievement between the two groups of teachers on the pre-test. However, on the post-test students of teachers not certified in math performed significantly worse in both math and algebra. Hawk et. al. concluded that students of certified math teachers achieve at a higher level compared to the students of teachers who do not hold math certificates.

In the state-level analysis of NAEP data, Darling-Hammond (1999) found that the relationship between the percent of teachers with full certification and a major in the field that they teach was positively and significantly associated with student achievement in math and reading in grade 4 and math in grade 8 (reading outcomes were not analyzed in grade 8), with standard estimates ranging from .64 to .87. The relationship between the percent of new teachers not certified in their field and student achievement was not significant. The analysis included two years of data and controlled for student characteristics and classroom size. Darling-Hammond recognized that the size of the relationship found in a state-level analysis might not be representative of the school-level or classroom-level relationship. Nonetheless,
she concluded that students could benefit academically from fully certified teachers who teach within their area of expertise.

Goldhaber and Brewer used data from the National Educational Longitudinal Study of 1988 (NELS) in a series of analyses that examined the effectiveness of teachers’ content area training (Goldhaber and Brewer, 1996; Goldhaber and Brewer, 1997; Goldhaber and Brewer, 2000). NELS is a national survey of approximately 24,000 8th grade students. A subset of students was surveyed again in the 10th grade and in the 12th grade. The survey links student and teacher records and provides information about teacher training and certification. In the 1996 study, Goldhaber and Brewer analyzed student achievement in the 10th grade math, science, English, and history. The results showed a positive relationship between content-area certification and student achievement in both math and science, with the estimates of 2.17 in math and 1.13 in science. In English and history, the estimates were not statistically significant. The analysis controlled for the 8th grade achievement, family background, and additional teacher characteristics. The researchers hypothesized that positive results in math and science could be due to the general ability of math and science teachers and not their subject-specific training. To test this theory, they analyzed the effect of math and science certification on achievement in English and history (the sample included teachers certified in math or science who were assigned to teach English or history). If content training in math and science served only as a proxy to general ability, then the researchers expected that the results would show a positive impact of math and
science certification on student achievement in English and history. This was not the case: neither math nor science certification was significantly related to achievement in English or history. The researchers concluded that in technical subjects such as math and science, subject specific training of teachers was an important element of student success. In 1997, Goldhaber and Brewer focused on student achievement in 10th grade math only. The results were similar to the 1996 study: certification in math was associated with a 2-point increase in math achievement. When researchers included teacher behavior variables into the model (e.g., time devoted to small groups, using oral questions, doing administrative tasks), the relationship was reduced to 1.4 points, but remained statistically significant. In both the 1996 and the 1997 studies Goldhaber and Brewer also analyzed teacher effect based on the certificate status only (i.e., fully certified teachers vs. uncertified teachers). They found a negative relationship between full certification and student achievement in math and science. This finding highlights the fact that not accounting for teachers’ content-area expertise may lead to different conclusions about the effect of teacher training on student outcomes.

Returning to the same data in 2000, Goldhaber and Brewer analyzed 12th graders achievement in math and science. The researchers differentiated between teachers holding standard content-area certificates, probationary content-area certificates,4 emergency content-area certificates,5 and private school certificates,5 and

---

4 Probationary certificates are issued to new teachers who met all of certification requirements but have not completed a probationary teaching period.
non-certified teachers (this group included teachers who were not certified at all and teachers who were certified, but teaching out-of-field). The results showed that, controlling for the prior achievement and school, family, and classroom characteristics, content-area certification had a positive effect on student learning. In math, students of teachers with content-area certification experienced approximately 1.3-point gain in test scores relative to students of non-certified teachers and teachers with private school certificates. The effects were smaller in science, but followed the same pattern. The researchers found no evidence that teachers with standard content-area certificates outperformed teachers with emergency content-area certificates. Darling-Hammond et al. (2001) criticized this finding because of a small subsample of teachers (24 science and 34 math teachers of the overall sample of 3,469 math and science teachers) and because of the methodological issues (e.g., not using sample weights, multicollinearity of variables). Darling-Hammond et al. also pointed out that teachers with the emergency certificates in the field that they teach are often experienced career teachers who await completion of requirements for a standard license, which makes their background similar to the background of teachers with standard certificates.

Using teacher and student records from the New York City, Boyd, Loeb, Wyckoff, Lankford, and Rockoff (2007) found limited support for the content area certification. The researchers constructed a dataset that linked teachers to students

---

5 Teachers with private school certificates likely have not come into the teaching profession via a traditional teacher preparation program.
and included approximately 65,000 to 80,000 student records per grade from 2000 to 2005. They estimated separate models for grades 4-5 and 6-8 with math and English Language Arts as student outcomes. In both subjects, content area certification was not significantly associated with gains in student achievement, although the coefficients were positive for math in grades 4-5 (the coefficients for grades 6-8 were not reported). Boyd et al. explained this result by high correlations between the variables included in the analysis. According to the researchers, when all measures were included into the model, the importance of individual qualifications (e.g., content area certification) was underestimated. They did not, however, report whether analyses with individual measures of teacher quality had been attempted.

**College Background as a Measure of Teacher Quality**

Although certificates are perhaps a more comprehensive measure of teacher quality, the one that combines preparation in content area and in education, sometimes researchers used college records to measure teacher training, often separating content area training and education training. Content area and education training are, however, the two core elements of certification, making the studies that used these measures comparable to the ones that used certification and therefore relevant to the present review. An example of such study is the one conducted by Monk (1994). Using data from the Longitudinal Study of American Youth (LSAY), Monk measured teacher training by their undergraduate and graduate college coursework and majors. LSAY was administered to high school students, teachers,
and parents and includes math and science achievement tests conducted in 1987, 1988, and 1989. The base sample for the study consisted of 2,829 students. Monk found a positive relationship between teachers’ undergraduate math courses (5 or fewer) and gains in math achievement in grades 10 and 11. He found no link between a math major and achievement. The relationship between teachers’ undergraduate and graduate coursework in math education and math achievement was positive in both grades with the majority of coefficients significant. The test of difference between the impact of content courses and education courses revealed that education courses had a significantly stronger impact on math achievement. The results of subject-matter preparation in science were mixed, with some estimates positive and some negative. The impact of science major was equally unclear, with non-significant estimates for the 10th grade and negative estimates for the 11th grade. The undergraduate educational coursework in science appeared to benefit students in the 11th grade, while graduate coursework was positively associated with the 10th graders’ achievement. In a follow-up study, Monk and King (1994) focused on content area training only. They found that teacher training had a differential impact on students of different ability: highly achieving 10th graders had higher gains in 10th grade if their teacher took more math courses, but for the low achieving 10th graders’ the effect did not appear until the next school year. Science results of the follow-up study were again perplexing, with some estimates showing a positive effect of the content courses and some estimates showing a negative effect. Monk suggested that using a composite test score as a dependent
variable for the science outcomes could have been a crude measure for capturing the effect of teacher training. He also thought that the differences in course taking in grades 10 and 11 could have been another reason for the mixed findings. Nonetheless, the researchers concluded that the evidence presented in both studies was encouraging, suggesting that teacher preparation in education was as important to student achievement as teacher preparation in a content area. It should be noted, however, that all positive relationships reported in both studies were small, with the coefficients ranging from .05 to .17.

In a cross-sectional analysis of NELS, Chaney (1995) studied the effectiveness of teachers taking advanced courses in math. He found that math test scores of the 8th graders whose teachers took advanced math courses were approximately 1.12-points higher than the scores of their peers whose teachers did not take advanced math. In science, students whose teaches had 40 or more credits in physical and life science also scored approximately 1-point higher than their peers. The relationship between teachers’ education courses and student achievement was not significant in both math and science. However, in a descriptive analysis of means Chaney found that students whose teachers took a combination of advanced mathematics courses and mathematics education courses had the highest average scores. Goldhaber and Brewer, who also examined NELS in a series of analyses in 1996, 1997, and 2000, found a significant and positive relationship between student achievement in 10th and 12th grade math and teachers’ bachelor’s degrees in math. The relationship between teachers’ degrees in science and student achievement in science was
positive and significant only for the 10th graders. The estimates were relatively small, suggesting that students whose teacher had a bachelor's in math or science scored approximately 0.7-points higher than their peers. In English and history, the relationship between teachers' degrees and student achievement was not significant. Rowan et. al. (1997) replicated the positive relationship between teachers' degrees and student achievement in math found by Goldhaber and Brewer. Using multilevel modeling and controlling for the prior achievement, student and school characteristics, the researchers found a positive link between teachers' undergraduate or graduate degrees in math and math achievement of 10th graders. Although significant, the effect was once again small at .015 standard deviations.

Using NAEP files, Greenberg et al. (2004) found that 8th graders scored nearly 5-points higher in math if their teachers held a major or minor in math or math education. The relationship between a major or minor in education and math achievement was not significant. The analysis controlled for school and student characteristics, but not for the prior achievement. Wenglinsky (2000) applied multilevel structural equation modeling to the same data and found that teachers' majors or minors in math and science, including math and science education, were positively associated with the 8th grade test scores in both subjects. Wenglinsky estimated that students whose teachers held degrees in math or science were 39% ahead of students whose teachers did not have relevant degrees. When he controlled for student characteristics, classroom characteristics, and teacher
practices, the relationship between teachers’ majors and student outcomes was modest with the standard estimate of .09.

Aaronson, Barrow, and Sander (2007) relied on a different approach to study the effect of teacher characteristics. Using results of the Test of Achievement and Proficiency from '96-'97 to '98-'99, the researchers first estimated teachers’ overall ability to teach math. Next, they regressed a variety of factors, including teachers’ college background, on the overall ability to teach math. The results suggested a weak link between teachers’ bachelor degrees in math and their ability to teach math with a standard estimate of .003. The study used test scores of the 9th grade CPS students.

So far this review has focused on the studies that were published since 1994. The research of teacher effectiveness dates further back and it is best described using the studies that relied on meta-analysis of publications released prior to 1994. Generally, the results of these studies suggest that teacher training in the subject matter and in education is beneficial to student achievement. For example, Druva and Anderson (1983) analyzed 65 studies conducted between 1966 and 1975 that investigated the relationship between science teachers’ characteristics and either their behavior in the classroom or student outcomes. They found a positive correlation between student achievement in science and the number of science courses teachers took. They also found a positive correlation between student achievement in biology and the number of biology courses taken by their teachers. Ashton and Crocker (1987) reported a positive relationship between the number of
credits in education and student outcomes in four out of seven studies. The relationship between the number of credits in a subject area and student outcomes was positive in five out of fourteen studies. The researchers concluded that training in education was as important to student achievement as training in a subject matter. Hanushek examined the relationship between school characteristics, including teacher preparation, and student outcomes in a series of analyses. In the most recent paper based on 90 studies published before 1994, he found that teacher test scores (measured by an achievement or IQ test) were most consistently related to higher student outcomes. The relationship between teacher education and student outcomes was less consistent, with fewer studies reporting a positive relationship, particularly when prior student achievement was included in the analysis (Hanushek, 1997). Hedges, Laine, and Greenwald (1994) and Greenwald, Hedges, and Laine (1996) re-analyzed the majority of studies selected by Hanushek using a different statistical approach. They found a more consistent positive relationship between teacher education and student outcomes. They also showed that a negative relationship between teacher education and student outcomes that appeared in some of the analyses was influenced by only one of the selected studies. Neither Hanushek nor Hedges et al. or Greenwald et al. have examined teacher training with regard to the subject area that they taught. The reports were also unclear on the specific elements of teacher background selected to measure teacher education.
Research Limitations and Contributions of Current Study

The main focus of this literature review has been on two attributes of teacher quality: expertise in content area and in education, measured using certification or college background. It is reasonable to anticipate that teachers who know their subject and who know how to teach it will be successful in the classroom. The evidence presented in this paper supports this idea: researchers usually find that teachers with adequate training in a subject area and in education have a positive effect on student learning. Other literature reviews offer a similar conclusion (Bolyard & Moyer-Packenham, 2008; Darling-Hammond, 1999; DeAngelis, Presley, & White, 2005; Rice, 2003; Wayne & Youns, 2003).

While the majority of papers included in the current review suggest a positive impact of teacher training, they share a number of shortcomings. These shortcomings, briefly stated in the introduction, include the studies’ primary focus on math and science as well as the focus on high school. The focus on math and science is not surprising, considering that math and science are core subjects deemed essential to students’ success in college and future careers and, on a larger scale, to the nations’ competitiveness on the world stage (U. S. Department of Education, 2008). However, the researchers devoted considerably less attention to other subjects that may be equally important to students’ success, such as reading. The emphasis of prior research on high school is also not surprising given the need for high school teachers prepared to teach advanced courses to college-bound students. The U.S. Department of Education, for instance, recommends that college-
bound high school students take 3 to 4 years of math, including Algebra II, Geometry, Trigonometry, and Calculus, and 2 to 4 years of science, including Biology, Chemistry, Earth Science, and Physics (U.S. Department of Education, 2009; USAA Educational Foundation, 2010). However, middle grades are also important for college preparation. In fact, the U. S. Department of Education describes middle grades as a starting point for college readiness and recommends that students begin taking advanced courses, such as Algebra-I, in 7th grade, and Algebra-II in 8th grade (U. S. Department of Education, 2010). It seems reasonable to expect that much like high school teachers, middle school teachers should have adequate training to be able to teach advanced content. Unfortunately, unlike high school studies, middle grades studies find limited support to the idea that advanced training of middle grades teachers is beneficial to student learning. For example, recent NAEP results showed that 8th graders who were taught by teachers with undergraduate math major scored on average 9-points higher than their peers. Yet a state-level analysis revealed that in some states there was no difference in math outcomes between students taught by math majors and students taught by teachers without a math major (Cavanagh, 2009). A few middle grades studies described earlier in this review present some evidence that it is beneficial to student learning to have middle grades teachers who are experts in the subject that they teach and know how to teach it (Chaney, 1995; Darling-Hammond, 1999; Greenberg et al., 2004; Hawk et al., 1986; Wenglinsky, 2000). However, these studies have many methodological differences between them.
Mixed findings in the area of middle grades research could be the result of a small number of studies that focused on middle grades (U. S. Department of Education, 2008; Williams et al. 2010). Inconsistencies in the middle grades teaching structure could also be at play. For instance, in some schools middle grades are taught by teachers in a self-contained classroom, where the same teacher teaches all subjects. Other schools have a departmentalized structure where students receive instruction in different subjects from teachers who only teach these subjects. A combination of the two is also possible. A variety of teaching structures increases the likelihood of complications and mistakes in the reporting of teaching assignments. For example, schools may report all or most of their teachers as self-contained, when in reality some of the teachers teach a single subject in a departmentalized setting. Errors in reporting lead to flawed data, making it difficult for the researchers to study teacher effect. Although only a few researchers specifically describe having problems identifying teaching assignments in middle grades data (Clotfelter et al., 2007; Darling-Hammond et al., 2005; Fetler, 1999; Laczko-Kerr & Berliner, 2002), the deficit of middle grades studies suggests that it may be a common issue.

The goal of the current study is to expand what is currently known about the effectiveness of teacher training in middle grades. It does that by examining the relationship between student outcomes and teacher training in math and reading at one of the largest public school districts in the Midwest. The district was selected for the study, because it requires high standards for the training of middle grades
teachers. The training standards have been formalized in a district policy. In addition, the district was selected because it serves a large population of minority and low-income students who have been the focus of much work and research around improving student achievement.

The study’s primary outcome – the nature of the relationship between teacher training and student achievement in middle grades – is examined using a multilevel model of change. Using a multilevel model of change allows for a longitudinal analysis of the relationship, which expands on the analytical methods used in earlier research. The study addresses the following research questions:

*Question 1:* What is the pattern of change in teacher training and student outcomes in middle grades math and reading at the district under study?

*Question 2:* What is the nature of longitudinal relationship between school-level outcomes in middle grades’ math and reading and teacher training in math and reading in the district under study?
CHAPTER THREE

METHOD

The goal of the study is to examine teacher training and student outcomes in math and reading at one of the largest public school districts in the Midwest. Design of the study is observational research (Graziano & Raulin, 2004). The study used administrative and assessment data collected and archived by the district under study and by the State Board of Education where the district is located. All analyses were conducted using SAS 9.2 (SAS Institute, 2012).

Sample

The study used administrative and assessment public school records for five school years, 2007 through 2011. Within that time, the district operated a total of 490 elementary schools. Of these, 414 schools were included in the study. The remaining 76 schools were not included, because they did not serve middle grades or served a limited number of middle grades (e.g., served grades PK through 5 or K through 6). All teacher and student records included in the study were nested within the 414 public schools included in the study.

The study used records of students in grades 3 through 8 who took a standardized achievement test in math or reading. In a given year, an average of 146,248 students in grades 3 through 8 took one or a combination of standardized tests, with the average of 74,025 students in grades 6 through 8. The study also used
records of full-time elementary school teachers hired by the district between the school years of 2007 through 2011, with the average number of teachers at 12,992. Table 1 shows counts of records included in the study and the number of schools each year of analysis.

Table 1: Records included in the study

<table>
<thead>
<tr>
<th></th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of schools</td>
<td>401</td>
<td>405</td>
<td>406</td>
<td>398</td>
<td>400</td>
</tr>
<tr>
<td>Test records grades 3-8</td>
<td>141,586</td>
<td>155,130</td>
<td>150,399</td>
<td>144,732</td>
<td>139,394</td>
</tr>
<tr>
<td>Test records grades 6-8</td>
<td>77,761</td>
<td>78,528</td>
<td>75,138</td>
<td>70,667</td>
<td>68,031</td>
</tr>
<tr>
<td>Teacher records</td>
<td>13,606</td>
<td>13,725</td>
<td>13,042</td>
<td>12,328</td>
<td>11,907</td>
</tr>
</tbody>
</table>

Data and Instrumentation

Data

The study used five sources of data. Four sources were provided by the State Board of Education where the district under study is located and include employment records, endorsement records, standardized test scores, and school records. An additional source of data were the district's school records posted on a public website.

Employment Records The file provided individual employment records of public school teachers hired by the district between 2007 and 2011. Records specified teachers’ schools, position descriptions (e.g., elementary teacher, junior high/middle teacher), and teaching assignments (e.g., elementary self-contained, algebra, Language Arts).
Endorsement Records The file provided information on individual teachers’ endorsements, date of issue, and grade level of endorsements. Examples of endorsements include *general science, mathematics, reading teacher*. Examples of grade level descriptions include *middle school, kindergarten to 4th grade, junior high school*.

Standardized Test Scores The file provided individual students’ scaled scores in math, science, and reading for all students who took one or a combination of standardized tests between 2007 and 2011. Additional information included students’ schools and grades. Math and reading test scores were available for grades 3 through 8. The file also included scaled science scores for grades 4 and 7. Science scores were not included in the study, because the science test used by the district was designed to assess scientific thinking and critical thinking skills instead of the actual content knowledge, making it unsuitable for measuring the impact of teacher training.

School Records School records were available at the grade- and/or school-level in the files provided by the State Board of Education and in the files posted online by the district. The records included information about the schools’ names, locations, and student population size. The records also included information about student demographics such as race, attendance, percent of low-income students, percent of students with limited English proficiency, and percent of students receiving free or reduced-priced lunch. Information about average class size, and time spent on math and reading was also available for the study. Some school
records, for example attendance, were available at both the grade and the school levels. Other records, for example mobility and dropout rates, were only available at the school level. Appendix A provides the description of all variables that were available for the study.

**Instrumentation**

The study used administrative and assessment data to measure middle grades outcomes (dependent variable), teacher training (independent variable), and student and school characteristics (control variables).

**Student Outcomes** Student scaled scores on standardized math and science tests were used to measure student outcomes. The tests were administered in the spring semester of each school year to all students in grades 3 through 8. As the dependent variable, the study used the schools’ yearly averages of test scores for grades 6 through 8, which are considered middle grades in the district under study. In order to capture student growth, the scores on standardized tests are vertically scaled, allowing students in higher grades to earn higher scores (for a description of vertical scaling see the University of Iowa, 2013). To ensure that the average test scores calculated for the study had not been disproportionately affected by the number of students in higher grades, the simple arithmetic yearly averages were compared to the weighted averages via paired t-tests. All results were not significant, with the p-values in math ranging from .16 to .96 and in reading from .10 to .85, leading to the decision to use arithmetic averages as the study's dependent variable.
Both math and reading test scores had five waves of data from 2007 to 2011. The possible ranges of scores for the 2011 school year were 120 to 410 in math and 120 to 369 in reading. The ranges were similar or identical for the previous school years.

**Teacher Training** To identify teachers with advanced training, the study used endorsements. In the district under study, endorsements are issued as part of the certification process to identify content area specialists. For example, a middle grades math endorsement identifies a teacher who has adequate training in math and in middle grades pedagogy and who, under the district’s policy, could teach math in grades 6 through 8. Endorsement requirements are identical within each content area and have not changed during the years included in the study, making endorsements a suitable measure of advanced teacher training both cross-sectionally and longitudinally. The study’s independent variable was the school-level count of teachers with middle grades math or reading endorsements who were also assigned to teach math or reading in grades 6 through 8. Teachers with middle grades endorsements in math or reading who were not assigned to teach in their field and/or those who were not assigned to teach middle grades were not included in the counts.

**Student and School Characteristics** School records were used to measure relevant student and school characteristics, such as attendance and dropout rates, percent of low-income students, instructional differences between schools, etc.
Student and school characteristics were available at the grade and/or school level. Appendix A provides a description of all variables and their level of aggregation.

**Procedure**

The State Board of Education where the district under study is located provided the majority of records used in the study. The files were provided under the data sharing guidelines established by the Family Educational Rights and Privacy Act, i.e., FERPA (U.S. Department of Education, 2012) and did not require a special research agreement.

Administrative data provided by the Board included employment and endorsement records of individual teachers. Although the data were provided at the individual level, the records had been encrypted with a fabricated ID, which allowed for the linking across employment and endorsement files, but prevented the identification of individual teachers. Assessment records provided by the Board included individual students’ scaled test scores, state school numbers, and grades. Student records did not have a student-level ID and could not be linked across years. Students could only be linked to schools and grades within a given year. Records could not be used to identify individual students. A state school number was the only common element between the teacher records and assessment records.

The study used data posted by the district under study on a public website to extract student and school characteristics. The information was available at the grade and/or school-level, preventing the identification of students.
Prior to the beginning of analysis, the researcher combined information from different data sources into one dataset. All data were aggregated to the middle-grades and/or school-level to allow merging across teacher and student files. The study’s dependent and independent variables were aggregated to the middle grades level. To facilitate multilevel analysis, the final dataset was organized to have one line of data for each year for each school, up to five rows for the years 2007 through 2011 (modeled after person-period format described in more detail by Singer & Willett, 2003). The final dataset contained a total of 2010 records for 414 schools.

**Analysis**

Data analysis was conducted in two stages: exploratory analysis and multilevel growth analysis. The goals of the exploratory analysis were to describe the sample and to examine patterns of student outcomes and teachers training independently. The primary goal of the multilevel growth analysis was to examine the longitudinal relationship between advanced teacher training and student outcomes in middle grades.

*Exploratory Analysis*

In the exploratory stage of analysis, I used descriptive statistics to summarize sample characteristics, such as student and school demographics. I examined the means of student achievement and teacher training in order to begin to understand the longitudinal nature of both. I used one-way repeated-measures analysis of variance (ANOVA) to analyze the significance of change in teacher training and student outcomes over time. Repeated-measures ANOVA is a suitable statistical
analysis of change when observations are not independent of each other (Howell, 2001), as was the case in the current study with teacher training and student outcomes measured repeatedly within-schools. As the first step in understanding the relationship between teacher training and student outcomes, I used Pearson product-moment correlation analysis (Howell, 2001). I used results of the exploratory analysis to address the first research question: what is the pattern of change in teacher training and student outcomes in middle grades math and reading at the district under study?

Multilevel Growth Analysis

The primary goal of multilevel growth analysis was to address the second research question: what is the nature of the longitudinal relationship between school-level outcomes in middle grades’ math and reading and teacher training in math and reading? The data available for the study fit well within the requirements for a longitudinal analysis outlined by Singer and Willet (2003): there were five waves of data; each wave of data was associated with a reliable time metric - a school year; the dependent variable - scaled test scores - was a continuous measure with outcomes comparable overtime.

I selected the multilevel growth analysis, because it is intended for the analysis of nested data. This was a desirable characteristic for the current study, because it used middle grades outcomes nested within schools. A multilevel growth model can be specified to analyze multiple waves of observations nested within a larger unit. In the current study, five waves of middle grades observations were
nested within individual schools. In addition, the data included time-varying and
time-invariant predictors that could be accommodated within a multilevel growth
model. Lastly, I selected the multilevel approach, because it permits missing
observations, as was the case with some schools included in the study¹ (Singer &
Willett, 2003; Raudenbush and Bryk, 2002).

The complete analysis included four models with the repeated middle grades
outcomes nested within schools. Middle grades outcomes were an aggregate of
outcomes for grade 6 through 8. Middle grades outcomes had five waves of
observations, 2007 through 2011. One set of models was used to examine math
outcomes and one set was used to examine reading outcomes. The first model in
each set was a base model that examined temporal changes in student achievement.
The second model examined temporal relationship between achievement and
teacher training independent of other predictors. To allow for the effect of teacher
training to vary over time, the third model included an interaction between time and
teacher outcomes. The final model controlled for the influence of time-varying and
time-invariant student and school characteristics. The analysis was conducted
following the steps outlined by Singer and Willett (2003), with considerations given

Multilevel analysis began with the examination of a base model of change in
middle grades outcomes:

\[
Y_{it} = \pi_{0i} + \pi_{1i}TIME_{it} + e_{it}
\]

¹ Fourteen schools that opened or closed between 2007 and 2011 had fewer than 5 observations.
Level-2: \( \pi_{0i} = \gamma_{00} + \varsigma_{0i} \)
\( \pi_{1i} = \gamma_{10} + \varsigma_{1i} \)

In this and the following models, \( Y_{ti} \) represents the dependent variable, achievement in middle grades math or reading, within an individual school, \( i \), at time \( t \); \( \pi_{0i} \) is the initial achievement level at \( TIME=0 \), which corresponds to the year 2007; \( \pi_{1i} \) is the achievement growth rate within the school \( i \), and \( e_{ti} \) is the error term for school \( i \). Level-2 equations represent averages across schools, with \( \gamma_{00} \) representing the average initial outcome and \( \gamma_{10} \) representing the average growth rate. Level-2 residuals, \( \varsigma_{0i} \) and \( \varsigma_{1i} \), represent deviation of within-school parameters from the overall averages of the initial status and change.

The two Level-2 parameters for the average initial status and average growth, \( \gamma_{00} \) and \( \gamma_{10} \), represent the model’s fixed effects. The model’s random effects, represented by \( e_{ti}, \varsigma_{0i} \) and \( \varsigma_{1i} \), allow for the parameters to vary within- and between-schools. Substituting Level-2 equations into the Level-1 equation produces the following composite specification of the base model, with the first set of brackets identifying fixed effects and the second set of brackets identifying random effects:

\[
Y_{ti} = [\gamma_{00} + \gamma_{10} \, TIME_{ti}] + [\varsigma_{0i} + \varsigma_{1i} + e_{ti}]
\]

Specifications for the models that follow are limited to composite specifications, because of the simplified interpretation and the ease of translating composite specifications into the appropriate SAS syntax. Samples of SAS syntax used to conduct multilevel analysis are presented in the Appendixes B and C.
The main interest in running the base models was to understand the magnitude and the direction of change in middle grades outcomes, while allowing for the random variation between-schools. The models residuals, $e_{0i}, \zeta_{0i}$ and $\zeta_{1i}$, were of less interest than their population variance components, $\sigma_{e}, \sigma_{0}$, and $\sigma_{1}$. In particular, values for $\sigma_{0}$, and $\sigma_{1}$ were used to determine the need of exploring the effect of time-invariant school-level predictors.

The next model included the time-varying independent variable, teacher training in middle grades math or reading. Teacher training was included as a fixed and a random effect:

$$Y_{ti} = \left[ \gamma_{00} + \gamma_{10} TIME_{ti} + \gamma_{20} ENDORS_{ti} \right] + \left[ \zeta_{0i} + \zeta_{1i} + \zeta_{2i} + e_{ti} \right]$$

This model presumes that the value of $Y$ (achievement in middle grades) at school $i$ at time $t$ depends on the school year ($TIME$) and teacher training ($ENDORS$), plus the residuals. Teacher counts were not transformed prior to inclusion into the model, because statistical tests do not prohibit from the use of count variables in the same way that continuous variables are used (McDonald, 2009). The counts were also not centered, because the variable had a meaningful value of zero, indicating schools that did not have any endorsed teachers. Adding independent variable into the model changed the interpretation of initial status, $\gamma_{00}$, into the initial achievement at schools with zero endorsed teachers. The interpretation of growth in achievement, $\gamma_{10}$, became growth in achievement controlling for the effect of teacher training.
A particular interest in running this model was to examine the main effect of teacher training, $\gamma_{20}$, and to compare the growth estimate, $\gamma_{10}$, between the base model and the expanded model. A change in the growth estimate would indicate that accounting for teacher training helped explain some of the observed changes in middle grades outcomes. A significant main effect of training would suggest that changes in achievement occurred in concert with changes in teacher training.

To allow for the effect of teacher training to vary over time, the next model included a cross-product of time and training as a fixed and random effect:

$$Y_{ti} = [\gamma_{00} + \gamma_{10}\text{TIME}_{ti} + \gamma_{20}\text{ENDORST}_{ti} + \gamma_{30}(\text{ENDORST}_{ti} \times \text{TIME}_{ti})]$$

$$+ [\varsigma_{0i} + \varsigma_{1i} + \varsigma_{2i} + \varsigma_{3i} + \epsilon_{ti}]$$

A significant main effect of interaction, $\gamma_{30}$, would suggest that teacher training impacted student outcomes differently over time. A change in the growth estimate, $\gamma_{10}$, would indicate that the addition of interaction helped explain some of the change in achievement. With the addition of interaction, the interpretation of the main effect of teacher training, $\gamma_{20}$, changed to indicate the difference between the average initial status of schools with and without endorsed teachers. In running this model, I also examined the possibility of quadratic and cubic relationship between teacher training and student achievement. The estimates were not significant and therefore were not included in this and the final model.

The final model controlled for the effect of time-varying and time-invariant student and school characteristics (in the level-1/level-2 specification, time-varying variables would be the level-1 variables, and time-invariant variables would be the
level-2 variables). All middle grades control variables were time-varying (COV1), and all school-level control variables were time-invariant (COV2). In the theoretical model below, subscript $t$ indicates the time-varying nature of middle-grades variables:

$$Y_{ti} = [\gamma_{00} + \gamma_{10} \text{TIME}_{ti} + \gamma_{20} \text{ENDORS}_{ti} + \gamma_{30} (\text{ENDORS}_{ti} \times \text{TIME}_{ti}) + \gamma_{40} \text{COV1}_{ti} + \gamma_{01} \text{COV2}_i + \gamma_{11} (\text{COV2}_i \times \text{TIME}_{ti}) + \gamma_{21} (\text{COV2}_i \times \text{ENDORS}_{ti}) + \gamma_{31} (\text{COV2}_i \times \text{ENDORS}_{ti} \times \text{TIME}_{ti}) + \gamma_{41} (\text{COV2}_i \times \text{COV1}_i)] + [\zeta_{0i} + \zeta_{1i} + \zeta_{2i} + \zeta_{3i} + \zeta_{4i} + \epsilon_{ti}]$$

The final model only included predictors with significant estimates that helped explain variance components and improved the goodness-of-fit statistics. To minimize multicollinearity, I did not include variables that were highly correlated. All control variables were grand-mean centered. All proportion variables were log-transformed prior to centering, because log transformation produced a better improvement in the distribution of values compared to other transformations.²

² The additional transformations were the Arcsine and square root transformations (Tabachnick & Fidell, 2007)
CHAPTER FOUR

RESULTS

This chapter describes results of the analysis. Sample characteristics are described first. Teacher training and student achievement in middle grades math and reading are described next. The last section summarizes findings of the multilevel growth analysis used to examine the longitudinal relationship between teacher training and student outcomes.

Sample Characteristics

Table 2 presents average sample characteristics across middle grades and schools. As shown in the table, from 2007 to 2011 the largest student demographic in the sample was Black and the second largest demographic was Hispanic. The combined percent of White and Asian students was approximately 12% each year.\(^1\) The majority of students were eligible for free or reduced priced lunch. Approximately 12% of students had limited English proficiency and approximately 12% of students were receiving special education services each year. Over time student attendance remained high, while chronic truancy rate remained low. Student mobility (i.e., students transferring between schools) fluctuated slightly, but

\(^1\) Additional races were not included in the study, because they made up 1% or less of student population.
Table 2: Sample characteristics across middle grades and schools

<table>
<thead>
<tr>
<th></th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mid</td>
<td>All</td>
<td>Mid</td>
<td>All</td>
<td>Mid</td>
</tr>
<tr>
<td></td>
<td>Grades</td>
<td>Schools</td>
<td>Grades</td>
<td>Schools</td>
<td>Grades</td>
</tr>
<tr>
<td>% of white students</td>
<td>8 (16)</td>
<td>9 (16)</td>
<td>8 (16)</td>
<td>9 (16)</td>
<td>9 (17)</td>
</tr>
<tr>
<td>% of black students</td>
<td>56 (42)</td>
<td>55 (43)</td>
<td>56 (42)</td>
<td>55 (43)</td>
<td>57 (43)</td>
</tr>
<tr>
<td>% of Hispanic students</td>
<td>30 (35)</td>
<td>30 (35)</td>
<td>30 (35)</td>
<td>30 (35)</td>
<td>31 (36)</td>
</tr>
<tr>
<td>% of Asian students</td>
<td>3 (8)</td>
<td>3 (8)</td>
<td>3 (8)</td>
<td>3 (8)</td>
<td>3 (9)</td>
</tr>
<tr>
<td>% of students with limited English</td>
<td>-</td>
<td>13 (15)</td>
<td>-</td>
<td>13 (16)</td>
<td>-</td>
</tr>
<tr>
<td>% of special education students</td>
<td>-</td>
<td>11 (4)</td>
<td>-</td>
<td>12 (5)</td>
<td>-</td>
</tr>
<tr>
<td>% of student receiving free or reduced lunch</td>
<td>-</td>
<td>79 (19)</td>
<td>-</td>
<td>84 (21)</td>
<td>-</td>
</tr>
<tr>
<td>Student attendance rate</td>
<td>-</td>
<td>94 (2)</td>
<td>-</td>
<td>94 (2)</td>
<td>-</td>
</tr>
<tr>
<td>Student mobility rate</td>
<td>-</td>
<td>24 (14)</td>
<td>-</td>
<td>27 (16)</td>
<td>-</td>
</tr>
<tr>
<td>Student truancy rate</td>
<td>-</td>
<td>2 (5)</td>
<td>-</td>
<td>3 (5)</td>
<td>-</td>
</tr>
<tr>
<td>% of teachers with emergency certificates</td>
<td>-</td>
<td>2 (3)</td>
<td>-</td>
<td>1 (2)</td>
<td>-</td>
</tr>
<tr>
<td>Minutes per day math is taught</td>
<td>55 (5)</td>
<td>55 (6)</td>
<td>54 (4)</td>
<td>54 (4)</td>
<td>55 (6)</td>
</tr>
<tr>
<td>Class size</td>
<td>25 (5)</td>
<td>25 (5)</td>
<td>26 (5)</td>
<td>26 (5)</td>
<td>27 (4)</td>
</tr>
<tr>
<td>Number of teachers</td>
<td>-</td>
<td>34 (15)</td>
<td>-</td>
<td>34 (15)</td>
<td>-</td>
</tr>
<tr>
<td>Mid grades enrollment</td>
<td>100 (100)</td>
<td>95 (103)</td>
<td>87 (103)</td>
<td>80 (102)</td>
<td>173 (97)</td>
</tr>
<tr>
<td>School enrollment</td>
<td>-</td>
<td>12 (312)</td>
<td>-</td>
<td>96 (319)</td>
<td>-</td>
</tr>
<tr>
<td>% Male students</td>
<td>-</td>
<td>51 (7)</td>
<td>-</td>
<td>51 (7)</td>
<td>-</td>
</tr>
</tbody>
</table>
on average remained approximately 23% each year. The distribution of male and female students remained nearly equal each year of the analysis.

With the exception of one year, the schools included in the study employed 1% or less of teachers with emergency certificates. The average number of teachers per school was approximately 30. The average size of middle grades class was approximately 25 students each year. Middle grades math was taught approximately 55 minutes per day (teaching time for reading was not available). The average middle grades enrollment was highest in 2007 and continued dropping slightly each subsequent year, similarly to the overall school enrollment.

Racial makeup of student body was the only characteristic available at both the middle grades and school levels. For each year of the analysis and each racial group, one-sample $t$-tests were used to compare middle grades to school-level averages (UCLA, 2013). All comparisons were not significant, with $p$-values ranging from .21 to .98, suggesting that the distribution of races in schools included in the analysis remained similar between middle grders and the entire school population over time.

Overall, descriptive statistics suggest little fluctuation from year to year in sample characteristics across middle grades and schools. Since the majority of elementary schools were included in the sample, its characteristics are likely reflective of the entire district.
Exploratory Analysis of Student Achievement and Teacher Training

Student Achievement

The analysis of student achievement began with the examination of means in math and reading scaled scores overtime. The results show that middle grades math and reading scores have been increasing each year since 2007. The test score means are summarized in Table 3 and illustrated graphically in Figure 1.

Table 3: Middle grades test scores over time

<table>
<thead>
<tr>
<th>Year</th>
<th>Math scores (M, SD)</th>
<th>Reading scores (M, SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>248 (14)</td>
<td>233 (11)</td>
</tr>
<tr>
<td>2008</td>
<td>249 (14)</td>
<td>235 (11)</td>
</tr>
<tr>
<td>2009</td>
<td>251 (14)</td>
<td>235 (11)</td>
</tr>
<tr>
<td>2010</td>
<td>253 (14)</td>
<td>236 (11)</td>
</tr>
<tr>
<td>2011</td>
<td>256 (14)</td>
<td>238 (11)</td>
</tr>
</tbody>
</table>

Figure 1: Middle grades test scores over time

The statistical significance of change in student test scores over time was examined using one-way repeated-measures ANOVA. The results, summarized in
Table 4, suggest that the change in middle grades math outcomes was significant, $F(4, 1548) = 210.60, p < .0001$. The change in reading outcomes was also significant, $F(4, 1548) = 167.29, p < .0001$. The results remained significant applying the Greenhouse-Geisser and Huynh-Feldt corrections used when the equality of variance assumption is violated, which was the case in this analysis according to Mauchly's test of sphericity for math ($\chi^2=1450.5, p < .0001$) and reading ($\chi^2=1004.7, p < .0001$).

### Table 4: Results of one-way repeated measures ANOVA for test scores

<table>
<thead>
<tr>
<th>Effect</th>
<th>MS</th>
<th>df</th>
<th>$F$</th>
<th>$p$</th>
<th>Greenhouse-Geisser</th>
<th>Huynh-Feldt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math test scores</td>
<td>Time</td>
<td>3476.54</td>
<td>4</td>
<td>210.60</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>16.51</td>
<td>1548</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading test scores</td>
<td>Time</td>
<td>1405.50</td>
<td>4</td>
<td>167.29</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>8.4</td>
<td>1548</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Teacher Training**

The examination of teacher training began with the analysis of the distribution of endorsed and non-endorsed teachers across schools. Table 5 shows that with the exception of 2009 the percent of schools that had one or more math-endorsed teacher teaching within her field had increased. For example, in 2007 approximately 51% of elementary schools had at least one math-endorsed teacher teaching within her field and by 2011 86% of elementary schools had at least one math-endorsed teacher teaching within her field. The number of math-endorsed teachers at a single school also increased: in 2007 18% of schools had two or more math-endorsed teachers, and in 2011 50% of schools had two or more math-
endorsed teachers. However, with the exception of 2001 the percent of schools with more than two math-endorsed teachers remained relatively low. The percent of schools with one or more reading-endorsed teacher has been increasing until 2009, but dropped since to the level observed in 2007, or approximately 33%. The percent of schools that had two or more reading-endorsed teachers fluctuated slightly each year, but remained relatively low at or around 10%. Table 5 also shows that at schools that employed math- or reading-endorsed teachers there was little variability in the endorsed teacher counts within a given year and over time, with the majority of schools employing one or two endorsed teachers.

Table 5: Across-school distribution of endorsed teachers teaching in-field

<table>
<thead>
<tr>
<th></th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 endorsed teachers</td>
<td>191 (48)</td>
<td>115 (28)</td>
<td>139 (34)</td>
<td>90 (23)</td>
<td>58 (15)</td>
</tr>
<tr>
<td>1 endorsed teacher</td>
<td>134 (33)</td>
<td>157 (39)</td>
<td>155 (28)</td>
<td>152 (38)</td>
<td>144 (36)</td>
</tr>
<tr>
<td>2 endorsed teachers</td>
<td>55 (13)</td>
<td>88 (22)</td>
<td>77 (19)</td>
<td>90 (23)</td>
<td>103 (26)</td>
</tr>
<tr>
<td>&gt; 2 endorsed teachers</td>
<td>21 (5)</td>
<td>45 (11)</td>
<td>35 (9)</td>
<td>66 (16)</td>
<td>95 (24)</td>
</tr>
<tr>
<td>Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 endorsed teachers</td>
<td>271 (68)</td>
<td>271 (67)</td>
<td>239 (59)</td>
<td>248 (62)</td>
<td>271 (68)</td>
</tr>
<tr>
<td>1 endorsed teacher</td>
<td>113 (28)</td>
<td>92 (23)</td>
<td>108 (27)</td>
<td>108 (27)</td>
<td>102 (26)</td>
</tr>
<tr>
<td>2 endorsed teachers</td>
<td>15 (4)</td>
<td>32 (8)</td>
<td>40 (10)</td>
<td>36 (9)</td>
<td>21 (5)</td>
</tr>
<tr>
<td>&gt; 2 endorsed teachers</td>
<td>2 (.5)</td>
<td>10 (2)</td>
<td>19 (5)</td>
<td>6 (2)</td>
<td>6 (2)</td>
</tr>
</tbody>
</table>

The total yearly counts of endorsed teachers are summarized in Table 6. For every year of the analysis, the district had a higher number of in-field teachers with math endorsements than in-field teachers with reading endorsements. The count of
teachers with math endorsements dropped in 2009, but increased substantially in 2010 and 2011. The count of teachers with reading endorsements fluctuated: increased until 2009, but dropped since. Graphically, teacher counts are presented in Figure 2.

Table 6: Yearly counts of in-field teachers with middle grades endorsements

<table>
<thead>
<tr>
<th></th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td># of teachers with math endorsements</td>
<td>312</td>
<td>486</td>
<td>427</td>
<td>570</td>
<td>714</td>
</tr>
<tr>
<td># of teachers with reading endorsements</td>
<td>149</td>
<td>190</td>
<td>248</td>
<td>199</td>
<td>163</td>
</tr>
</tbody>
</table>

Figure 2: Yearly counts of in-field teachers with middle grades endorsements

One-way repeated-measures ANOVA was used to analyze the significance of change in teacher counts. The results, summarized in Table 7, suggest a significant change in the count of teachers with math endorsements, $F (4, 1548) = 83.74$, $p<.0001$. The results also suggest a significant change in the count of teachers with
reading endorsement, \( F(4, 1548) = 10.70, p < .0001 \). While the result in reading is significant, the value of \( F \) is substantially smaller than the value of \( F \) in math, which was likely a function of smaller number of reading-endorsed teachers that also fluctuated over time. The results for both math and reading remained significant applying the Greenhouse-Geisser and Huynh-Feldt corrections used when the equality of variance assumption is violated, which was the case in this analysis according to Mauchly's test of sphericity for teacher counts in both math (\( \chi^2 = 964.2, p < .0001 \)) and reading (\( \chi^2 = 538.3, p < .0001 \)).

Table 7: Results of one-way repeated-measures ANOVA for endorsements

<table>
<thead>
<tr>
<th>Effect</th>
<th>MS</th>
<th>df</th>
<th>( F )</th>
<th>( p )</th>
<th>Greenhouse-Geisser</th>
<th>Huynh-Feldt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math endorsements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>56.98</td>
<td>4</td>
<td>83.74</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>.68</td>
<td>1548</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading endorsements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>3.59</td>
<td>4</td>
<td>10.70</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>.34</td>
<td>1548</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Relationship Between Achievement, Training, and Other Study Variables*

A Pearson product-moment correlation analysis was used as the first step in exploring the relationship between student outcomes and teacher training in middle grades. The analysis used five-year school averages of test scores and endorsed teacher counts. The correlation between math scores and the count of teachers with math endorsements was significant and positive, \( r = .14, p < .01 \), suggesting a higher math achievement at the schools with a higher number of math-endorsed teachers. According to Cohen’s convention, the correlation effect size is small (Cohen, 1988).
The correlation between reading scores and the count of teachers with reading endorsements was significant and negative, $r=-.10$, $p<.05$, suggesting a lower reading achievement at the schools with a higher number of reading-endorsed teachers. Although this initial finding is perplexing, it could have been the result of the fluctuating number of teachers with reading endorsements revealed in the analysis of means. The correlation effect size is small (Cohen, 1988). Table 8 summarizes correlation results for both math and reading.

**Table 8: Results of Pearson product-moment correlation analysis between mean test scores and endorsements**

<table>
<thead>
<tr>
<th></th>
<th>Math scores</th>
<th>Reading scores</th>
<th>Math endorsements</th>
<th>Reading endorsements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math scores</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading scores</td>
<td>0.95***</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math endorsements</td>
<td>0.14**</td>
<td>0.09</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Reading endorsements</td>
<td>-0.06</td>
<td>-0.10*</td>
<td>0.30***</td>
<td>1</td>
</tr>
</tbody>
</table>

*p$<.05; **p$<.01; ***p$<.001

Pearson product-moment correlation analysis was also used to gain an initial understanding of the relationship between all variables available for the study. Summarized in Table 9, the results show a significant correlation of medium to large size between math and reading scores and racial composition of the student population. Other student characteristics, such as the percent of students receiving free or reduced priced lunch, attendance and mobility rates, and chronic truancy, were also highly correlated with the test scores. Of the remaining variables, class size, number of teachers, and the middle grades and school enrollment were significantly and positively correlated with the test scores, although with the smaller
Table 9: Results of Pearson product-moment correlation analysis between average test scores and control variables

<table>
<thead>
<tr>
<th></th>
<th>Math scores</th>
<th>Reading scores</th>
<th>% of white students</th>
<th>% of black students</th>
<th>% of white students</th>
<th>% of Hispanic students</th>
<th>% of students with limited English</th>
<th>% of special education students</th>
<th>% of students receiving free or reduced lunch</th>
<th>Student attendance rate</th>
<th>Student mobility rate</th>
<th>Student truancy rate</th>
<th>Class size</th>
<th>% of teachers with emergency certificates</th>
<th>Minutes per day math is taught</th>
<th>Number of teachers</th>
<th>Mid grades enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math scores</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading scores</td>
<td>.95***</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of white students</td>
<td>.75***</td>
<td>.76***</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of black students</td>
<td>-.49***</td>
<td>-.42***</td>
<td>-.62***</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of Hispanic students</td>
<td>.68***</td>
<td>.67***</td>
<td>.79***</td>
<td>-.41***</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of students with limited English</td>
<td>.34***</td>
<td>.26***</td>
<td>.55***</td>
<td>-.80***</td>
<td>.45***</td>
<td>.95***</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of special education students</td>
<td>-.12*</td>
<td>-.12*</td>
<td>.08</td>
<td>.09</td>
<td>.01</td>
<td>.02</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of students receiving free or reduced lunch</td>
<td>-.74***</td>
<td>-.79***</td>
<td>-.67***</td>
<td>.24***</td>
<td>-.52***</td>
<td>-.16***</td>
<td>-.04</td>
<td>-.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student attendance rate</td>
<td>.75***</td>
<td>.72***</td>
<td>.61***</td>
<td>-.64***</td>
<td>.53***</td>
<td>.60***</td>
<td>.57***</td>
<td>-.21***</td>
<td>-.42***</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student mobility rate</td>
<td>-.79***</td>
<td>-.81***</td>
<td>-.61***</td>
<td>.46***</td>
<td>-.48***</td>
<td>-.41***</td>
<td>-.31***</td>
<td>.17***</td>
<td>.61***</td>
<td>-.75***</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student truancy rate</td>
<td>-.68***</td>
<td>-.68***</td>
<td>-.57***</td>
<td>.50***</td>
<td>-.47***</td>
<td>-.46***</td>
<td>-.43***</td>
<td>.17***</td>
<td>.48***</td>
<td>-.81***</td>
<td>.67***</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class size</td>
<td>.32***</td>
<td>.31***</td>
<td>.029***</td>
<td>-.37***</td>
<td>.18***</td>
<td>.33***</td>
<td>.32***</td>
<td>-.09</td>
<td>-.17***</td>
<td>.35***</td>
<td>-.31***</td>
<td>-.26***</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of teachers with emergency certificates</td>
<td>-.10*</td>
<td>-.12*</td>
<td>-.06</td>
<td>-.07</td>
<td>-.04</td>
<td>.12**</td>
<td>.15**</td>
<td>-.06</td>
<td>.08</td>
<td>-.02</td>
<td>.13**</td>
<td>.04</td>
<td>-.08</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minutes per day math is taught</td>
<td>-.06</td>
<td>-.07</td>
<td>-.04</td>
<td>.03</td>
<td>-.04</td>
<td>-.04</td>
<td>-.02</td>
<td>.04</td>
<td>-.03</td>
<td>-.01</td>
<td>.07</td>
<td>-.08</td>
<td>-.01</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of teachers</td>
<td>.21***</td>
<td>.14**</td>
<td>.23**</td>
<td>-.49**</td>
<td>.20***</td>
<td>.47***</td>
<td>.49***</td>
<td>-.21***</td>
<td>.04</td>
<td>.35***</td>
<td>-.23***</td>
<td>-.26***</td>
<td>.41***</td>
<td>-.01</td>
<td>-.06</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Mid grades enrollment</td>
<td>.18***</td>
<td>.13**</td>
<td>.16**</td>
<td>-.45***</td>
<td>.12**</td>
<td>.39***</td>
<td>.41***</td>
<td>-.29***</td>
<td>.06</td>
<td>.33***</td>
<td>-.21***</td>
<td>-.28***</td>
<td>.45***</td>
<td>-.04</td>
<td>-.06</td>
<td>.93***</td>
<td>1</td>
</tr>
<tr>
<td>School enrollment</td>
<td>.20***</td>
<td>.14**</td>
<td>.21**</td>
<td>-.50**</td>
<td>.16***</td>
<td>.45***</td>
<td>.47***</td>
<td>-.31***</td>
<td>.04</td>
<td>.35***</td>
<td>-.22***</td>
<td>-.26***</td>
<td>.44***</td>
<td>&lt;.001</td>
<td>-.06</td>
<td>.98***</td>
<td>.96***</td>
</tr>
</tbody>
</table>

*p<.05; **p<.01; ***p<.001
effect sizes. Many of the school and student characteristic variables were highly correlated with each other. For example, the percent of White student was highly correlated with the additional races and the truancy rate was moderately to highly correlated with nearly all variables in the study. These findings suggested the wisdom of selectivity when including variables in the multilevel analysis in order to avoid multicollinearity.

**Multilevel Growth Analysis**

A multilevel growth analysis was used to examine the longitudinal relationship between teacher training and student outcomes. Results of the analysis are summarized below, with math outcomes described first, followed by the reading outcomes.

**Math**

The analysis began with specifying the unconditional growth model in order to understand the magnitude and the direction of change in the math outcomes. Recall that in this and the following models, the outcome was the school-level aggregate of middle grades test scores with five waves of data for each year of the analysis. The results show that the average scaled score in middle grades math in 2007 was a non-zero score of 247.62 ($p<.001$). The results also show that on average math scores increased linearly at 1.87 points per year ($p<.001$). The results suggest a significant within-school variation in math outcomes over time, $\sigma_e=8.97$, $p<.001$. The results also suggest a significant between-school variation in the initial status,
\( \sigma_0 = 196.51, p<.001 \), and in the rate of growth, \( \sigma_1 = 3.16, p<.001 \). A significant between-school variation in the initial status and in the rate of growth suggest the usefulness of introducing school-level time-invariant predictors (level-2 predictors in the level-1/level-2 model specification). Results of the unconditional model and all subsequent models are summarized in Table 10.

The introduction of the main effect of teacher training in the next model had little impact on the estimate of change, with the parameter remaining essentially the same (Model 2 in Table 10). The main effect of teacher training was not significant, \( \gamma_{20} = .13, p =.24 \). The inclusion of teacher training into the model increased the variance estimates and did not improve the goodness-of-fit statistics, suggesting a poorer overall fit of this model to data.

The effect of teacher training on math outcomes could vary over time, which was accounted for in the next model by introducing an interaction between training and time (Model 3 in Table 10). The interaction effect was statistically significant, \( \gamma_{30} = .22, p<.001 \). The interaction was examined further by comparing the pattern of growth in math achievement between schools that had at least one endorsed teacher and school that had zero endorsed teachers. Figure 3 shows results of plotting the outcomes. The results suggest that the initial math achievement at schools that had at least one math-endorsed teacher was lower than at schools that didn’t have any endorsed teachers. However, math achievement at schools that had at least one math-endorsed teacher improved faster. By 2009, math achievement at
Table 10: Results of multilevel growth analysis in math (standard errors in parentheses)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(1) Unconditional model</th>
<th>(2) Including training</th>
<th>(3) Including training*time</th>
<th>(4) Full model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time-varying:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept (status in 2007)</td>
<td>$\gamma_{00}$</td>
<td>247.62***</td>
<td>247.52***</td>
<td>248.06***</td>
</tr>
<tr>
<td>Time (change in math achievement)</td>
<td>$\gamma_{01}$</td>
<td>(.10)</td>
<td>(.11)</td>
<td>(.14)</td>
</tr>
<tr>
<td>Teacher training in math</td>
<td>$\gamma_{02}$</td>
<td>.13</td>
<td>-.40*</td>
<td>-.29</td>
</tr>
<tr>
<td>Training*Time</td>
<td>$\gamma_{30}$</td>
<td>.22***</td>
<td>.21***</td>
<td>(.06)</td>
</tr>
<tr>
<td>% of White students in middle grades</td>
<td>$\gamma_{40}$</td>
<td>(.06)</td>
<td>.58**</td>
<td>(.21)</td>
</tr>
<tr>
<td>% of Asian students in middle grades</td>
<td>$\gamma_{50}$</td>
<td>(.26)</td>
<td>1.32***</td>
<td></td>
</tr>
<tr>
<td>Time-invariant:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of students receiving free/reduced lunch</td>
<td>$\gamma_{61}$</td>
<td>-.12.52***</td>
<td>(1.30)</td>
<td></td>
</tr>
<tr>
<td>Rate of student mobility</td>
<td>$\gamma_{62}$</td>
<td>-.8.94***</td>
<td>(.65)</td>
<td></td>
</tr>
<tr>
<td>% of students with limited English</td>
<td>$\gamma_{63}$</td>
<td>1.15***</td>
<td>(.22)</td>
<td>(.43**</td>
</tr>
<tr>
<td>Time*mobility</td>
<td>$\gamma_{11}$</td>
<td>(.15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time*students with limited English</td>
<td>$\gamma_{12}$</td>
<td>-.22***</td>
<td>(.06)</td>
<td></td>
</tr>
<tr>
<td>Variance Components</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within-school</td>
<td>$\alpha_6$</td>
<td>8.97***</td>
<td>9.02***</td>
<td>8.64***</td>
</tr>
<tr>
<td>In intercept</td>
<td>$\alpha_0$</td>
<td>196.51***</td>
<td>199.3***</td>
<td>197.87***</td>
</tr>
<tr>
<td>In rate of change</td>
<td>$\alpha_1$</td>
<td>3.16***</td>
<td>3.27***</td>
<td>3.88***</td>
</tr>
<tr>
<td>In teacher training</td>
<td>$\alpha_2$</td>
<td>.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In teachers*time</td>
<td>$\alpha_3$</td>
<td>.07*</td>
<td>.07*</td>
<td></td>
</tr>
<tr>
<td>% of White students in middle grades</td>
<td>$\alpha_4$</td>
<td>1.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of Asian students in middle grades</td>
<td>$\alpha_5$</td>
<td>1.21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Goodness-of-fit

<table>
<thead>
<tr>
<th></th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12606.2</td>
<td>12607.6</td>
</tr>
<tr>
<td></td>
<td>12578.3</td>
<td>12620.5</td>
</tr>
</tbody>
</table>

*p<.05; **p<.01; ***p<.001
schools with at least one math-endorsed teacher was similar to math achievement at schools with zero math-endorsed teachers. By 2010, schools with endorsed teachers surpassed schools without the endorsed teachers, a trend that continued into the next year.

Figure 3: Math achievement over time at schools with and without math-endorsed teachers

Parameter estimates in Model 3 help qualify the interaction further, showing that schools with one math-endorsed teacher grew .22 points faster compared to schools with zero endorsed teachers. By 2011, the average math achievement at schools with one endorsed teacher was approximately half a point higher than at the
schools with zero endorsed teachers, and achievement at schools with two endorsed teachers was nearly one point higher.¹

A comparison of within-school variance estimates between Models 3 and 2 suggests that the inclusion of interaction helped explain 4% of within-school variation in math outcomes ((9.02-8.64)/9.02). The estimates of between-school variation in the initial status and rate of growth changed slightly, which was expected given that school-level predictors had not been added into this model. The variance component of teacher training was not included in Model 3, because its estimate in the previous model did not indicate a significant between-school variation. The variance component of interaction was significant, \( \sigma_3=0.07, p<0.05 \), suggesting the usefulness of examining the impact of school-level predictors on the interaction slope. Including interaction in Model 3 improved the goodness-of-fit statistics, indicating a better overall fit of this model to data.

In order to further examine the relationship between teacher training and student outcomes and to continue improving the model’s fit, the final model (Model 4 in Table 10) controlled for impact of school and student characteristics. The model controlled for the time-varying effect of the percent of White and Asian students in middle grades. The model controlled for the impact on the initial status of the percent of students receiving free or reduced priced lunch, the rate of student

---

¹ Endorsed teachers=0, year=4: 248.06 + 1.57*4 - .40*0 +.22*4*0 = 254.34; Endorsed teachers=1, year=4: 248.06 + 1.57*4 -.40*1 +.22*4*1 = 254.82; Endorsed teachers=2, year=4: 248.06 + 1.57*4 -.40*4 +.22*4*1 = 255.30.
mobility, and the percent of students with limited English proficiency. The model also controlled for the impact of student mobility and the percent of students with limited English on the rate of change. All control variables were grand-mean centered. The equation for the final model in math was as follows:

\[ Y_{ti} = [\gamma_{00} + \gamma_{10} TIME_{ti} + \gamma_{20} ENDORS_{ti} + \gamma_{30} (ENDORS_{ti} \times TIME_{ti}) + \gamma_{40} WHITE_{ti} + \gamma_{50} ASIAN_{ti} + \gamma_{01} FRL_{i} + \gamma_{02} MOBIL_{i} + \gamma_{03} LEP_{i} + \gamma_{11} (MOBIL_{i} \times TIME_{ti}) + \gamma_{12} (LEP_{i} \times TIME_{ti})] \]

The results suggest a positive association over time between math achievement in middle grades and the percent of White students in middle grades, \( \gamma_{40}= .58, p<.01 \). The results also suggest a positive association between math achievement in middle grades and the percent of Asian students in middle grades, \( \gamma_{50}= 1.32, p<.001 \). Two school-level predictors (i.e., time-invariant school averages) were significantly and negatively associated with the initial math achievement, including the percent of students receiving free or reduced priced lunch, \( \gamma_{01}= -12.52, p<.001 \), and the rate of student mobility, \( \gamma_{02}= -8.94, p<.001 \). The percent of students with limited English proficiency was significantly and positively associated with the initial math achievement, \( \gamma_{03}= 1.15, p<.001 \). However, the percent of students with limited English was negatively associated with the growth in math outcomes, \( \gamma_{12}= -.22, p<.001 \). An increase in the rate of student mobility was positively associated with growth in math outcomes, \( \gamma_{11}= .43, p<.01 \). However, since student mobility was also strongly and negatively associated with the initial math achievement, schools
with high mobility in 2007 were likely at a disadvantage in math learning compared to schools with lower initial mobility rates, even given the possibility of faster growth.

The estimate of the initial math status in the final model was the estimate at the schools with zero endorsed teachers at the average value of all control variables. The estimate was significant and essentially unchanged compared to the previous models, $\gamma_0=248.04, p<.001$. The estimate of growth in math achievement also changed little and remained significant, $\gamma_1=1.55, p<.001$. The interaction effect between teacher training and time remained significant in the final model as well, $\gamma_3=.21, p<.001$. This finding suggests that even after accounting for the differences in student and school characteristics, math achievement at schools with endorsed teachers improved faster and by 2011 was approximately half a point higher at schools with one endorsed teacher (1.1 points higher at schools with two endorsed teachers, 1.65 points at schools with three endorsed teachers, etc.).

In running the final model, I examined different covariance matrixes to determine a within-school error structure that would result in the best model fit. The autoregressive structure, which presumes a stronger correlation between observations that are closer in time, produced the best fit and was used in the final model (see Appendix D for the results of comparing fit statistics using the autoregressive covariance structure and the compound symmetry structure used in SAS by default). The autoregressive within-school error structure was combined
with the unstructured specification for the variance-covariance matrix for intercepts and slopes (Littell, Milliken, Stroup, & Wolfinger, 1994; Singer, 1998; Uekawa, 2004). The additional parameters and autoregressive covariance structure did not improve the estimate of within-school variance. In fact, it increased from 8.64 in the previous model to 21.97. This estimate, however, is likely a better representation of correlation between the repeated school-level observations. The remaining variance estimates improved substantially. The final model helped explain 89\% of between-school variance in the initial status \(((197.87 - 22.12)/197.87)\), 53\% of between-school variance in the rate of change \(((3.88 - 1.82)/3.88)\), and 71\% of between-school variance in the cross-product of training and time \(((.07 -.02)/.07)\), with the remaining variance not significant. The goodness-of-fit statistics in the final model also improved, suggesting a better overall fit of this model to data.

*Reading*

Similarly to math, the analysis of longitudinal relationship between reading outcomes and teacher training in reading began with specifying the unconditional growth model. The results, summarized in Model 1 of Table 11, suggest that the average scaled score in middle grades reading in 2007 was a non-zero score of 232.93 \((p<.001)\). On average, reading scores increased at the significant rate of 1.14 points per year \((p<.001)\). The results also suggest a significant within-school variation in reading outcomes, \(\sigma_e = 6.08, p<.001\), a significant between-school variation in the initial status, \(\sigma_0 = 109.94, p<.001\), and a significant between-school
variation in the rate of growth, $\sigma_1 = 1.09, p < .001$. Significant between-school variance estimates suggest the usefulness of introducing school-level time-invariant predictors.

Model 2 in Table 11 shows results of adding the independent variable, teacher training. The main effect of training was not significant, $\gamma_{20} = .02, p = .84$. In this model, the estimate of the initial status in reading is the estimate of the initial status at school with zero endorsed teachers, and the estimate of growth controls for the number of endorsed teachers. Adding teacher training into the model did not appreciably change parameters of the initial status or the rate of growth. The estimate of between-school variance in teacher training was not significant, $\sigma_2 = .46, p = .10$, and the estimates of variance components first obtained in Model 1 remained essentially the same. Changes in the goodness-of-fit statistics were trivial.

To allow the effect of teacher training on reading outcomes to vary over time, the next model included a cross-product of training and time (Model 3 in Table 11). The interaction effect was not significant, $\gamma_{30} = -.07, p = .31$, and its random variance component was also not significant, $\sigma_3 < .01, p = .37$. Within-school variance in Model 3 increased slightly compared to the previous model, while all of the remaining between-school variance components remained essentially unchanged. The goodness-of-fit statistics in Model 3 were worse compared to the previous models, indicating a poorer fit of this model to data.
### Table 11: Results of multilevel growth analysis in reading (standard errors in parentheses)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(1) Unconditional model</th>
<th>(2) Including training</th>
<th>(3) Including training*time</th>
<th>(4) Full model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (status in 2007)</td>
<td>$\gamma_{00}$</td>
<td>232.93*** (.52)</td>
<td>232.91*** (.53)</td>
<td>232.84*** (.53)</td>
</tr>
<tr>
<td>Time (change in reading achievement)</td>
<td>$\gamma_{10}$</td>
<td>1.14*** (.07)</td>
<td>1.15*** (.07)</td>
<td>1.19*** (.07)</td>
</tr>
<tr>
<td>Teacher training in reading</td>
<td>$\gamma_{20}$</td>
<td>.02 (.13)</td>
<td>.22 (.21)</td>
<td>.23 (.20)</td>
</tr>
<tr>
<td>Training*Time</td>
<td>$\gamma_{30}$</td>
<td>-.09 (.09)</td>
<td>-.10 (.08)</td>
<td></td>
</tr>
<tr>
<td>% of White students in middle grades</td>
<td>$\gamma_{40}$</td>
<td>.33* (.15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of Asian students in middle grades</td>
<td>$\gamma_{50}$</td>
<td>1.19*** (.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of students receiving free/reduced lunch</td>
<td>$\gamma_{51}$</td>
<td>-10.66*** (.75)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate of student mobility</td>
<td>$\gamma_{92}$</td>
<td>-5.50*** (.44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of special education students</td>
<td>$\gamma_{93}$</td>
<td>-1.38* (.64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of chronic truants</td>
<td>$\gamma_{94}$</td>
<td>-2.23*** (.35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of white students * % of special education students</td>
<td>$\gamma_{41}$</td>
<td>-1.00** (.35)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Variance Components

<table>
<thead>
<tr>
<th>Component</th>
<th>$\sigma$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Within-school</td>
<td>$\sigma_e$</td>
<td>6.08***</td>
<td>5.95***</td>
<td>6.05***</td>
</tr>
<tr>
<td>In intercept</td>
<td>$\sigma_0$</td>
<td>109.94***</td>
<td>110.63***</td>
<td>110.29***</td>
</tr>
<tr>
<td>In rate of change</td>
<td>$\sigma_1$</td>
<td>1.09***</td>
<td>1.08***</td>
<td>1.10***</td>
</tr>
<tr>
<td>In teacher training</td>
<td>$\sigma_2$</td>
<td>.46</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>In teachers*time</td>
<td>$\sigma_3$</td>
<td>.01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>% of White students in middle grades</td>
<td>$\sigma_4$</td>
<td>-</td>
<td>.28</td>
<td></td>
</tr>
<tr>
<td>% of Asian students in middle grades</td>
<td>$\sigma_5$</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

### Goodness-of-fit

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>11593.5</td>
</tr>
<tr>
<td>BIC</td>
<td>11609.6</td>
</tr>
</tbody>
</table>

*$p<.05$; **$p<.01$; ***$p<.001$
The final reading model controlled for the influence of student and school characteristics (Model 4 in Table 11). It controlled for the time-varying effect of the presence of White and Asian students. The model controlled for the impact on the initial status of the percent of students receiving free or reduced priced lunch, student mobility rate, the percent of special education students, and the percent of chronic truants. The model also accounted for the impact of special education students on the slope for the percent of White students in middle grades. The random effects of teacher training and the cross-product of training and time were not included, because their variance estimates were not significant in the previous models. Equation for the final model in reading was as follows:

\[ Y_{ti} = \left[ \gamma_{00} + \gamma_{10} \text{TIME}_{ti} + \gamma_{20} \text{ENDORS}_{ti} + \gamma_{30} (\text{ENDORS}_{ti} \times \text{TIME}_{ti}) + \gamma_{40} \text{WHITE}_{ti} + \gamma_{50} \text{ASIAN}_{ti} + \gamma_{01} \text{FRL}_{ti} + \gamma_{02} \text{MOBIL}_{ti} + \gamma_{03} \text{SPED}_{ti} + \gamma_{04} \text{TRUANT}_{ti} + \gamma_{41} (\text{SPED}_{ti} \times \text{WHITE}_{ti}) \right] + \left[ \varsigma_{0i} + \varsigma_{1i} + \varsigma_{4i} + \epsilon_{ti} \right] \]

Controlling for the school and student characteristics in the final model did not change the nature of the longitudinal relationship between teacher training and student outcomes in reading, suggesting no connection between the two. With the addition of control variables the estimate of the initial reading achievement remained essentially unchanged at 232.68 points \((p<.001)\). The rate of yearly growth increased by .12 points to the total of 1.30 points \((p<.001)\). An increase in the percent of White students was positively associated with reading achievement, \(\gamma_{40}=.33, p<.05\), as was the increase in the percent of Asian students, \(\gamma_{50}=1.19, p<.001\).
A number of student characteristics were significantly and negatively associated with the initial achievement in reading: the percent of students receiving free or reduced priced lunch, $\gamma_{01}=-10.66, p<.001$, the rate of student mobility, $\gamma_{02}=-5.50, p<.001$, percent of special education students, $\gamma_{03}=-1.38, p<.05$, and percent of chronic truants, $\gamma_{04}=-2.23, p<.001$. Although the presence of White students was positively associated with the middle grades achievement in math, an increase in the percent of special education students had a negative impact on that relationship, $\gamma_{41}=-1.00, p<.01$. Covariates did not have a significant impact on the rate of growth in reading achievement and therefore were not included in the final model.

Similarly to math, the final model in reading combined the autoregressive within-school error structure with the unstructured specification for the variance-covariance matrix for intercepts and slopes. The model did not improve the estimate of within-school variance: it increased from 6.05 in the previous model to 11.70, which was likely a better estimate of correlation between repeated observations. Compared to the previous model, the final model helped explain 94% of between-school variance in the initial status ((110.29 - 6.65)/110.29) and 74% of between-school variance in the rate of change ((1.10 - .29)/1.10), with the remaining variance in the rate of change no longer significant. The estimate of between-school variance in the percent of white students was not significant, $\sigma_{t}=.28, p=.25$. The estimate of between-school variance in the percent of Asian students was too low for the model to produce a probability estimate, suggesting no significant between-school
variation. The goodness-of-fit statistics in the final model also improved, suggesting a better overall fit of this model to data.
CHAPTER FIVE

DISCUSSION

The study was conducted with the goal of increasing the knowledge about the effectiveness of advanced teacher training in middle grades. Teacher training was measured using content-area certificates, i.e., endorsements. Using secondary data from one of the largest districts in the Midwest, the study examined trends in teacher training and student achievement in math and reading, and the nature of the longitudinal relationship between the two. This chapter summarizes results of the study. It continues with the discussion of implications, and concludes with the overview of study’s limitations and directions for future research.

Overview of Results

The study’s primary analytic goal was to examine the nature of the relationship between teacher training and student outcomes in math and reading in middle grades. The overview of results is organized around each subject area. The overview also includes a discussion of local policy context and trends in teacher assignments relevant to the interpretation of study’s findings.

Math

The study’s results suggest that over time the district under study experienced a significant change in the number of in-field math-endorsed teachers.
In 2007, the total number of in-field math-endorsed teachers was 312 and 52% of elementary schools had at least one math-endorsed teacher. By 2011, the number of teachers was 714 and 86% of schools had at least one math-endorsed teacher. At the same time, student achievement in middle grades math also grew. Math achievement at schools where middle grades students did not receive math instruction from math-endorsed teachers grew at the significant rate of 1.55 points per year on average. The study’s results suggest a beneficial effect of having students receive math instruction from math-endorsed teachers, with the impact larger at schools with a higher number of math-endorsed teachers. For example, schools with one math-endorsed teacher grew .21 points faster each year compared to schools with zero endorsed teachers, and schools that employed two math-endorsed teachers grew .42 points faster. Although math achievement at schools with endorsed teachers was lower in 2007 compared to schools without the endorsed teachers, by 2009 the difference between the schools disappeared, and by 2011 schools with one endorsed teacher were .55 points ahead of schools with zero endorsed teachers, and schools with two endorsed teachers were 1.1 points ahead.

In considering the differential impact of the number of math-endorsed teachers, recall that the study’s outcome was the average of middle grades outcomes at a given school in a given year. It is likely that schools with more endorsed teachers had more students taught by endorsed teachers. A higher number of students taught by endorsed teachers could have led to a larger proportion of better
achieving students at a given school, which would have led to a higher middle grades math average. Therefore, the increase in the strength of relationship between teacher training and student outcomes at schools with a higher number of math-endorsed teachers was likely indicative of the proportion of students receiving math instruction from endorsed teachers: the higher the number of students being taught math by content-specialists, the higher the school’s middle grades average. Consequently, this finding suggests that within the same schools some middle grade students could have been disadvantaged if they were among the students who were not receiving math instruction from a content specialist.

Understanding that the study used middle grades averages is also important in considering the size of the estimate for teacher training, which is relatively small at .21 points per year on average at schools with one math-endorsed teacher. The middle grades averages that the estimates were based included students who were and were not exposed to math-endorsed teachers. Therefore, the study could be underestimating the impact of teacher training in math on individual students. Another consideration is the cumulative nature of teacher effect (Sanders & Rivers, 1996). Potentially, a middle grader taught by a math-endorsed teacher in consecutive years could be accumulating extra points on top of the prior year's accelerated achievement, in which case the teacher effect could be increasing each year.
Although prior research on the effectiveness of advanced teacher training in middle grades is scarce (U. S. Department of Education, 2008; Williams, Kirst, Haertel et al., 2010), results of a small number of studies located in preparation for the current study found a positive effect of teacher training in math as measured by teaching certificates or college courses (Chaney, 1995; Darling-Hammond, 1999; Greenberg et al., 2004; Hawk et al., 1986; Wenglinsky, 2000). While the current study used endorsements to measure teacher training, its findings are supportive of earlier research, because in the district under study the process of gaining an endorsement is imbedded within teacher certification and is based on college background. The current study also expands and strengthens earlier research in demonstrating that the positive effect of teacher training in math could persist over time.

Reading

During the time period included in the analysis, the number of teachers with reading endorsements assigned to teach reading in middle grades fluctuated, but overall remained relatively low compared to the number of math-endorsed teachers. In 2007, the number of in-field reading-endorsed teachers was 149 and 32% of elementary schools had at least one reading teacher with a reading endorsement. The number peaked in 2009 at 248, when 42% of schools had at least one in-field reading-endorsed teacher. By 2011, the number of teachers dropped to 163, with 33% of elementary schools with at least one reading-endorsed teacher. At
the same time, student achievement in middle grades reading kept improving. At the
schools where students did not receive reading instruction from reading-endorsed
teachers, the average increase in reading achievement was significant at 1.30 points
per year.

The study’s results suggest that over time middle graders did not benefit
from receiving reading instruction from teacher with reading endorsements. This
finding could not be compared to earlier research on the effectiveness of teachers in
middle grades, because it did not examine reading. However, the results of studies
that focused on elementary grades were often similar to the current study’s in that
they found no teacher effect in reading or smaller effect in reading compared to
math or science. For example, Nye, Konstantopoulos, and Hedges (2004) found a
smaller overall teacher effect in reading compared to math in kindergarten through
third grade. Teacher education in reading had no impact on achievement in any
grades. One explanation of why researchers do not find a significant effect of teacher
training in reading could be that reading is not contained to learning within the
classroom. While math learning is more specialized and typically happens inside the
classroom, reading skills can be affected by what happens at home, in interactions
with peers, through media, etc. External influences may be corroding teacher effect
in reading, making it difficult to detect statistically with the measures of student
outcomes currently available to researchers (e.g., Hill, Gormley, Adelstein, &
Willemin, 2012; Nye et al., 2004).
Another reason why the study did not find an impact of teacher training in reading could be the low number of reading-endorsed teachers used to estimate the effect of training, especially since the estimates were calculated at school level. To understand why the number of reading-endorsed teachers remained low over time and why this was not the case for math-endorsed teachers, it is necessary to consider the local policy context that likely contributed to these differences.

Local Policy Context

In 2009, the district under study made a commitment to higher standards for middle grades teachers by instituting a policy that required core subjects to be taught by endorsed teachers. Math was included in the policy as a core subject, but not reading. The policy required that slightly over a hundred positions be designated to teachers endorsed in core subjects by 2010. The policy was amended in 2011 to require that the number of positions allocated to teachers endorsed in core subjects be based on the number of students at each school serving middle grades.

As schools began to comply with the new policy, they could have started to allocate more in-field teaching positions to teachers with middle grades endorsements in core subjects and fewer in-field positions to teachers endorsed in non-core subjects. Teacher records available for the study allowed for the comparison of teacher assignments in math (included in the policy as a core subject) and reading (not included in the policy as a core subject). Table 12, which is an
expanded version of Table 6, shows that indeed, since the policy was implemented in 2009 the number of in-field assignments of math-endorsed teachers has been increasing. The opposite is true for reading: while the total count of teachers with reading endorsements has been increasing, the number of in-field assignments of reading-endorsed teachers has been dropping since 2009. Recall that the study estimated teacher effect for teachers who were assigned to teach in their field. Thus as in-field assignments of reading-endorsed teachers dropped, so did the numbers that the study used to estimate the effect of teacher training.

Table 12: Yearly counts of teachers with middle grades endorsements teaching in-field and out-of-field

<table>
<thead>
<tr>
<th></th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>In-field</td>
<td>All</td>
<td>In-field</td>
<td>All</td>
</tr>
<tr>
<td># of teachers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with math</td>
<td>495</td>
<td>312</td>
<td>183</td>
<td>651</td>
<td>486</td>
</tr>
<tr>
<td>endorsements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of teachers</td>
<td>471</td>
<td>149</td>
<td>322</td>
<td>505</td>
<td>190</td>
</tr>
<tr>
<td>with reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>endorsements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To illustrate this point further, Figure 4 shows the percent of teachers with math and reading endorsements assigned to teach in their field over time. While in 2009 the percentages were similar, in the following years the percent of in-field assignments of math-endorsed teachers increased, while the percent of in-field assignments of reading-endorsed teachers dropped substantially.
Figure 4: Percent of math- and reading-endorsed teachers assigned to teach in-field

Considering the local policy context helps understand the differences in the number of math- and reading-endorsed teachers observed in the current study and used to estimate the impact of teacher training. It also sheds a light on the differences in teaching assignments that likely reduced the study's ability to estimate the impact of teacher training in reading. When these differences are taken into account, the fact that the study did not find an effect of teacher training in reading becomes more illustrative of changes in the district policy landscape than suggestive of no benefit of teacher training in reading.

**Study’s Implications & Contributions**

In recent years, many school districts in the United States, including the largest districts, started to require advanced training from middle grades teachers (e.g., Chicago Public Schools, 2008; Chicago Public Schools, 2011; District of
Columbia Office of the State Superintendent of Education, 2009; New York State Education Department, 2011; State of California Commission on Teacher Credentialing, 2011). The idea driving the demand for higher standards is that teachers with advanced training would be better equipped to foster student learning. The study’s results suggest that strengthening requirements for teacher training in middle grades may be a beneficial strategy for math education. Positive results in math suggest that middle graders could benefit from learning other subjects from content specialists.

Common sense suggests that teachers with advanced training should teach within their area of expertise to have a positive impact on student achievement. However, teachers are often assigned to teach outside of their area, as the current study demonstrates particularly in the subject of reading, with 75% of reading-endorsed teachers misassigned in 2011. This finding is in line with prior research that also shows that teachers often get assigned to teach subjects in which they don’t have specialized training (Hill & Gruber, 2011; Ingersol, 1999; Robinson, 1985; Seastrom, Gruber, Henke, McGrath, & Cohen, 2002), with the likelihood of out-of-field assignments higher in middle school than in high school (Ingersol, 1999; Seastrom et al., 2002). Results of the current study combined with the earlier research suggest that for teacher training to have a desired impact on student achievement in middle grades, school districts could benefit from combining stricter training standards for teachers with a measure of school accountability aimed at
improving the practice of teaching assignments. A caveat introduced by the study is that when schools are held accountable for teaching assignments only in some subjects, other areas of learning may experience a drop in the number of in-field assignments, as was the case with reading in the current study. School districts could benefit from policies that balance the needs for trained teachers across different areas of learning.

On a larger scale, results of the study are relevant to the ongoing debate about the usefulness of teaching certificates as a measure of teacher quality. As suggested by the study, certification may only be as effective as its application at the school level. Currently, school administrators have little or no accountability to ensure a fit between teaching certificates and teaching assignments, with the decisions about assignments almost exclusively made by the principals (Feng, 2010; Ingersol, 1999; Loeb, Kalogrides, & Beteille, 2011; Robinson, 1985). As a result, many teachers get misassigned and are not positioned to deliver the best learning for which they were trained and certified. As the nation continues the effort to improve teacher quality, a step in the right direction may be to take a closer look at the existing practice of teaching assignments and consider a system of school accountability that would maximize teacher training.

In conclusion, the current study contributed to a small but growing body of research on the effectiveness of teacher training in middle grades. The study supported earlier findings in math and laid the initial groundwork for the
exploration of teacher effectiveness in middle grades reading. In using the multilevel model of growth, the study expanded on research methodology used in earlier research and provided initial evidence that the positive effect of teacher training in middle grades could persist over time.

**Limitations and Directions for Future Research**

The study’s findings should be considered with a number of limitations in mind. Perhaps the biggest shortcoming of the study is the school level of analysis. While analyzing school-level averages is useful in understanding the relationships between the variables, the estimates based on averages may be over- or underestimating the impact of teacher training on individual students. Due to FERPA limitations, data available for the current study could not be analyzed at the level of individual students. However, if requirements to access individual records change in the future, additional research would benefit from analyzing the link between teacher training and student outcomes with more precision.

Of the 414 schools included in the study, 14 had missing observations. Of the schools with missing observations, 2 were missing four observations and 12 were missing 2 or 3 observations. The schools with missing observations were the ones that opened or closed during the period included in the analysis. The multilevel growth model selected for the analysis permits missing observations (Singer & Willett, 2003). However, the reader should be aware of this limitation.
Using standardized test scores to measure student achievement is another limitation of the current study. Test scores may not be sensitive enough to detect teacher effect, particularly at the school-level of analysis. In addition, researchers have raised concerns about the reliability and validity of standardized test scores in measuring achievement (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009) and about changes in the administration and scoring of the tests that could complicate conclusions based on the test score results (Luppescu, Allensworth, Moore, de la Torre, & Murphy, 2011). While recognizing these limitations, the study used test scores because they were the sole measure of student achievement available at the time the study was conducted that allowed for cross-sectional and longitudinal comparisons. Future research would benefit from supplementing test scores with other measures of student learning.

In the current study, as in other education production studies, direction of causality between teacher characteristics and student achievement cannot be asserted with complete confidence (Nye et al., 2004). Student, school, and neighborhood characteristics that the study did not assess may influence the distribution of teachers at different schools and in different classrooms. For example, higher achieving schools may put more emphasis on hiring middle grades teachers with advanced training. School administrators may assign more effective teachers to teach advanced students. In either situation, teacher training would not have a direct causal link to student achievement, but instead serve as a proxy for
school factors underlying differences in achievement. To determine causality more effectively, future research would benefit from testing the entire causal chain from teachers to student outcomes, including what happens in the classroom. Differences in what happens in the classrooms of teachers with advanced training and teachers without the advanced training could signal more definitively whether content specialists teach more effectively.

While teaching assignments were not the main focus of the study, the findings suggested a possible link between assignments and teacher effectiveness. Although some researchers examined trends in teacher assignments and factors that influence assignments, employment, and evaluation of teachers (Clotfelter et al., 2007; Feng, 2010; Hill & Gruber, 2011; Ingersol, 1999; Jacob, 2010; Jacob & Lefgren, 2005; Loeb, Kalogrides, & Beteille, 2011; Robinson, 1985; Seastrom, et al., 2002), the area remains largely unexplored and could benefit from additional research, particularly the research focused on the relationship between assignments and student learning.

Lastly, in considering generalizability of the study's findings, researchers should note that the study used data from a large school district that serves a high percent of low-income and minority students. The generalizability of findings is limited to other large districts that serve a similar population of students. Since the analysis controlled for student and school characteristics, to some extent the results
may also be generalized to districts that are smaller and/or serve a different population of students.
APPENDIX A

DEFINITION OF VARIABLES
### Definition of Variables

#### Level 1: Dependent Variables (Middle Grades Variables)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Characteristics/Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math: scaled test score</td>
<td>Interval</td>
<td>Possible range 120 to 410</td>
</tr>
<tr>
<td>Science: scaled test score</td>
<td>Interval</td>
<td>Possible range 120 to 390</td>
</tr>
<tr>
<td>Reading: scaled test score</td>
<td>Interval</td>
<td>Possible range 120 to 369</td>
</tr>
</tbody>
</table>

#### Level 1: Independent Variables (Middle Grades Variables)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Characteristics/Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math endorsed middle grades teachers who teach math</td>
<td>Interval</td>
<td>Number; 0 indicates no teachers</td>
</tr>
<tr>
<td>Science endorsed middle grades teachers who teach science</td>
<td>Interval</td>
<td>Number; 0 indicates no teachers</td>
</tr>
<tr>
<td>Reading endorsed middle grades teachers who teach reading</td>
<td>Interval</td>
<td>Number; 0 indicates no teachers</td>
</tr>
</tbody>
</table>

#### Level 1: Covariates (Middle Grades Variables)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Characteristics/Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math: national percentile rank</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Science: national percentile rank</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Reading: national percentile rank</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Percent of white students</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Percent of black students</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Percent of Hispanic students</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Percent of Asian students</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Enrollment</td>
<td>Interval</td>
<td>Number of students enrolled in middle grades</td>
</tr>
<tr>
<td>Average class size</td>
<td>Interval</td>
<td>Number of students per class</td>
</tr>
<tr>
<td>Minutes per day math</td>
<td>Interval</td>
<td>Number of minutes, grades 6 and 8</td>
</tr>
<tr>
<td>Minutes per day science</td>
<td>Interval</td>
<td>Number of minutes, grades 6 and 8</td>
</tr>
</tbody>
</table>
### Level 2: Covariates (School-Level Variables)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Characteristics/Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math: scaled test score</td>
<td>Interval</td>
<td>Possible range 120 to 410</td>
</tr>
<tr>
<td>Science: scaled test score</td>
<td>Interval</td>
<td>Possible range 120 to 390</td>
</tr>
<tr>
<td>Reading: scaled test score</td>
<td>Interval</td>
<td>Possible range 120 to 369</td>
</tr>
<tr>
<td>Math endorsed teachers</td>
<td>Interval</td>
<td>Number of math endorsed teachers at a school</td>
</tr>
<tr>
<td>Science endorsed teachers</td>
<td>Interval</td>
<td>Number of science endorsed teachers at a school</td>
</tr>
<tr>
<td>Reading endorsed teachers</td>
<td>Interval</td>
<td>Number of reading endorsed teachers at a school</td>
</tr>
<tr>
<td>Math: national percentile rank</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Science: national percentile rank</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Reading: national percentile rank</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Percent of white students</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Percent of black students</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Percent of Hispanic students</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Percent of Asian students</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Enrollment</td>
<td>Interval</td>
<td>Number of students enrolled at a school</td>
</tr>
<tr>
<td>Number of teachers</td>
<td>Interval</td>
<td>Number of teachers at a school</td>
</tr>
<tr>
<td>Attendance rate</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Chronic truants rate</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Dropout rate</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Mobility rate</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Students receiving free or reduced lunch</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Student with disabilities</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Students of limited English proficiency</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
<tr>
<td>Teachers with emergency or provisional credential</td>
<td>Interval</td>
<td>Possible range 0 to 100%</td>
</tr>
</tbody>
</table>

#### Identifying Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Characteristics/Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>School year</td>
<td>Ordinal</td>
<td>0 to 4; 0=2007 and 4=2011</td>
</tr>
<tr>
<td>District unit number</td>
<td>Nominal</td>
<td>N/A</td>
</tr>
<tr>
<td>School name</td>
<td>Nominal</td>
<td>N/A</td>
</tr>
<tr>
<td>Grades served at a school</td>
<td>Nominal</td>
<td>Example: grades 3-8, 6-8.</td>
</tr>
</tbody>
</table>
APPENDIX B

SAMPLES OF SAS CODE FOR MULTILEVEL MODEL IN MATH
Samples of SAS Code for Multilevel Models in Math

**Unconditional model**

```sas
proc mixed data=isat covtest noclprint;
   class unit;
   model math_m=time/solution ddfm=bw ;
   random intercept time/subject=unit type=un;
run;
```

**Including teacher training and training x time**

```sas
proc mixed data=isat_c covtest noclprint;
   class unit;
   model math_m=time math_end_teach_m time*math_end_teach_m / solution ddfm=bw ;
   random intercept time time*math_end_teach_m /subject=unit type=un;
run;
```

**Full model using autoregressive covariance structure of within-person error**

*Variable ‘wave’ was added to separate the time of observations within-person, values 1 to 5.*

```sas
proc mixed data=isat_c covtest noclprint;
   class unit wave;
   model math_m=time math_end_teach_m time*math_end_teach_m white_pct_m_log asian_pct_m_log fri_pct_log mob_pct_log lep_pct_log time*mob_pct_log time*lep_pct_log/ solution ddfm=bw ;
   random intercept time time*math_end_teach_m white_pct_m_log/subject=unit
   repeated wave/type=ar(1) subject=unit r;
run;
```
APPENDIX C

SAS CODE FOR PLOTTING INTERACTION IN MATH
SAS Code for Plotting Interaction in Math

Set up a dataset with predicted values for each year

data result_plot;
   do end = 0 to 8 by 1;
      year=0;
      math_score=248.06 + 1.57*year -.40*end +.22*year*end;
      output;
   end;
   do end=0 to 8 by 1;
      year=1;
      math_score=248.06 + 1.57*year -.40*end +.22*year*end;
      output;
   end;
   do end=0 to 8 by 1;
      year=2;
      math_score=248.06 + 1.57*year -.40*end +.22*year*end;
      output;
   end;
   do end=0 to 8 by 1;
      year=3;
      math_score=248.06 + 1.57*year -.40*end +.22*year*end;
      output;
   end;
   do end=0 to 8 by 1;
      year=4;
      math_score=248.06 + 1.57*year -.40*end +.22*year*end;
      output;
   end;
run;

Group values by endorsements

data result_plot;
   length end_num $10;
   set result_plot;

   if end~0 then end_num="1 or more";
   else end_num="0";
run;
**Interaction plot**

ods graphics on/reset = all height=4.5in width=4.5in noborder;

proc sgplot data = result_plot ;
  pbspline y=math_score x=year/group=end_num name="s1" nomarkers;
  keylegend "s1" / across= 1 location=inside noborder position=bottomright;

  label math_score = 'math test scores';
  label year = 'year';
  xaxis min = 0 max = 4 values=(0 to 4 by 1);
  yaxis min = 240 max = 260;
run;
quit;
ods graphics off;
APPENDIX D

MODEL FIT UNDER DIFFERENT COVARIANCE STRUCTURES
Math: model fit under the compound symmetry and autoregressive within-error covariance structures

<table>
<thead>
<tr>
<th></th>
<th>AIC (smaller is better)</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS</td>
<td>11890.60</td>
<td>15</td>
</tr>
<tr>
<td>AR(1)</td>
<td>11801.90</td>
<td>16</td>
</tr>
<tr>
<td>Difference</td>
<td>88.70</td>
<td>-1</td>
</tr>
</tbody>
</table>

AR(1) covariance estimate = .63, SE = .08, p<.0001

Reading: model fit under the compound symmetry and autoregressive within-error covariance structures

<table>
<thead>
<tr>
<th></th>
<th>AIC (smaller is better)</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS</td>
<td>10774.30</td>
<td>14</td>
</tr>
<tr>
<td>AR(1)</td>
<td>10718.00</td>
<td>15</td>
</tr>
<tr>
<td>Difference</td>
<td>56.30</td>
<td>-1</td>
</tr>
</tbody>
</table>

AR(1) covariance estimate = .51, SE = .08, p<.0001

Under additional structures, the fit was poorer or the models did not converge.


VITA

Natalya’s studies began at the Linguistics University in Minsk, Belarus. After graduation with a Master’s degree in English and Spanish Linguistics, she developed a passion for cultural exploration, which prompted her to seek new realities by making a home in the United States. Professional experiences, the complexities of cultural assimilation, and subsequent psychological responses seeded Natalya’s fascination in societal and interpersonal relationships. This newfound interest in human interactions led her to graduate studies in Applied Social Psychology at the Loyola University, Chicago. As her interests matured, Natalya chose to pursue a doctorate in Research Methodology at Loyola. In April of 2013 she successfully defended her dissertation on the subject of teacher training and student outcomes in middle grades.

Throughout her studies and professional experiences, Natalya became captivated by the intertwined nature of education, family, and community, with a particular interest in education. She has explored these interests by serving as a member of internal research and evaluation group at Chicago Public Schools between 2006 and 2011. Later she joined Chapin Hall at the University of Chicago as an Associate Researcher, where she has been since contributing to research benefitting children, families, and communities.
DISSERTATION APPROVAL SHEET

The dissertation submitted by Natalya S. Gnedko has been read and approved by the following committee:

Meng-Jia Wu, PhD, Director
Associate Professor, School of Education
Loyola University Chicago

Terri Pigott, PhD
Professor, School of Education
Loyola University Chicago

Kelci Price, PhD
Director, Research and Evaluation
The Colorado Health Foundation

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the committee with reference to content and form.

The thesis is therefore accepted in partial fulfillment if the requirements for the degree of Doctor of Philosophy.

_________________________________   ___________________________________________
Date       Director’s Signature
The dissertation submitted by Natalya S. Gnedko has been read and approved by the following committee:

Meng-Jia Wu, PhD, Director
Associate Professor, School of Education
Loyola University Chicago

Terri Pigott, PhD
Professor, School of Education
Loyola University Chicago

Kelci Price, PhD
Director, Research and Evaluation
The Colorado Health Foundation

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the committee with reference to content and form.

The thesis is therefore accepted in partial fulfillment if the requirements for the degree of Doctor of Philosophy.

________________________________________      _____________________________
Date       Director’s Signature