GILBERT, CAROLYN ANNE, Ph.D. Capturing the Diversity of English Language Learners' Cultural and Linguistic Backgrounds and the Influence on Math and Reading Achievement. (2015)
Directed by Dr. John T. Willse. 103 pp.

English language learners (ELLs) are diverse individuals with various cultural and linguistic backgrounds. Unlike native English speakers, they do not all share a common language. Additionally, ELL students were not all born in the same country and most have not spent the same number of years attending U.S. schools. ELL students are a heterogeneous group, but the current research does not sufficiently appreciate and recognize those differences. Examining the diversity of ELL students, by incorporating contextual variables with limited sample sizes, was accomplished by using a hierarchical linear modeling approach. The results showed that students classified as LEP in fourth grade demonstrated lower initial mean scores in both math and reading than did exited LEP students. Students in both the Asian and Austro-Asiatic native language groups demonstrated higher math and reading scores at initial status compared to students in the Spanish native language group. The number of years attending U.S. schools impacted math and reading. Students born in the U.S. or Canada demonstrated higher math scores in 4^{th} grade than did students in the Mexico/Central America/South America/Caribbean country of birth group.

These findings have implications for both future research and practice in terms of methodological choices and database management to emphasize and address the academic needs of ELL students.

CAPTURING THE DIVERSITY OF ENGLISH LANGUAGE LEARNERS' CULTURAL AND LINGUISTIC BACKGROUNDS AND THE INFLUENCE ON MATH AND READING ACHIEVEMENT

by

Carolyn Anne Gilbert

A Dissertation Submitted to
the Faculty of the Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2015

Approved by

____
Committee Chair

To Dr. Gongshu Zhang

APPROVAL PAGE

This dissertation written by Carolyn Anne Gilbert has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair ______________________
John Willse

Committee Members ______________________
Micheline Chalhoub-Deville

____
Ye He

____
Robert Henson

____
Date of Acceptance by Committee

____
Date of Final Oral Examination

ACKNOWLEDGMENTS

I am indebted to John Willse, my advisor and dissertation chair, for his encouragement and support from my beginning days in the ERM program to this final stage. Thank you for your matter-of-factness, gentle prodding, and willingness to see me through. At times things seemed impossible and endless, but you never gave up on me.

I would not have been drawn to this area of study if not for Micheline Chalhoub-Deville. Participating in your courses on Language Assessment and Testing was an impactful experience that informed and changed my thinking. Your knowledgeable, passionate approach, delivered by your lilting voice, had a tremendous influence on me. Thank you for your guidance along this journey.

I am grateful to Bob Henson for his technical assistance in HLM. The detailed, written comments that you so kindly provided for my proposal defense were immensely helpful and directed me well this last year. Thank you as well for so clearly explaining concepts and offering advice when I had questions. The final thank you to my committee goes to Jane He. I have a stronger understanding about long-term English learners because of your expertise and suggestions. Thank you for your kindness and willingness to help me whenever I needed.

A different thank you goes to my friend, Fang Chen. We began the program together, and you supported me throughout our many shared courses. I will always think of you fondly and am so glad you were a major part of this experience.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Within their framework for equitable assessment policies for English language learners (ELLs), LaCelle-Peterson and Rivera (1994) adamantly stated:

> The diversity among ELLs needs to be recognized, lest all ELLs be regarded as a monolithic group with a single defining educational characteristic: use of a non-English language. Indeed, while language represents an important, educationally significant variable that is most often conspicuous by its absence from U.S. educational discourse, it is only one of many educationally relevant characteristics of an individual English language learner, whose whole identity…must be taken into consideration in educational decisions (p. 59-60).

Research Problem

When discussing ELL students, many educators refer to ELLs as one similar group, as if having a native language other than English is the single contribution to their academic performance. This narrow thinking obscures important variations in the students' cultural background, language skills, and academic preparation that they bring to school each day. ELLs are diverse individuals with various cultural and linguistic backgrounds. Unlike native English speakers, they do not all share a common language. Additionally, ELL students were not all born in the same country and most have not spent the same number of years attending U.S. schools. Many older children arrive in the U.S. with little or no formal schooling in their native language (Freeman & Freeman, 2002). Although most ELLs were born in the United States, they were raised in homes where no one was English fluent (Garcia, 2000). Examining this varied demographic context should yield distinctive information about the academic achievement of ELL students.

Much of the educational research on ELL students has compared the academic performance of ELL and never-ELL elementary-age students and focused on the achievement gap that persists between the two groups (Abedi & Lord, 2001; Abedi, 2002; Durán, 2008). These obvious findings are not remarkable, and have done little to inform what precipitates this difference or what could be improved to positively impact achievement. While the academic differences between ELL and never-ELL students have been traditionally emphasized, adjusting the lens to focus on variability within groups has been suggested (Genesee, Lindholm-Leary, Saunders, & Christian, 2005; Kim & Herman, 2009; Wolf, Herman, & Dietel, 2010). ELL students are a heterogeneous group, but the current research does not sufficiently appreciate and recognize those differences.

Of course, one language does not describe all ELL students, yet most of the research addresses ELLs as one group. Availability and use of data have presented challenges. In some cases, the data available for study have created obstacles for researchers to form appropriate native language groups, use relevant sample sizes, or perform strong statistical analyses. Spanish speakers were oversampled in the limited research studies that did discuss the relationship between native language and academic achievement of ELL students (Hofstetter, 2003; Kieffer, 2008; Robinson, 2008). The use of data mainly from Spanish speakers could be attributed to the vast number of students in most sampling pools with Spanish as their first language. Many researchers acknowledged the limitations they experienced when creating sample groups and moved forward with only a Spanish language group (Hofstetter, 2003; Hopstock & Stephenson, 2003; Robinson, 2008). Studying the academic achievement of ELL students as one group or only as Spanish speakers fails to recognize the complexity and diversity of students' linguistic backgrounds. Moreover, heterogeneity is present within individual languages; for example, Spanish speakers living in the U.S. were born in different countries and have varying cultural backgrounds, socioeconomic status, and academic experiences.

Recent research involving the use of hierarchical linear modeling approaches has delivered more informative results. A few researchers still traveled the traditional route by comparing academic performance of ELL and never-ELL students (Lesaux, Rupp, & Siegel, 2007; Kieffer, 2008; Roberts, Mohammed, & Vaughn, 2010), but their strong research designs did afford improved estimates. Several studies substantively demonstrated progress in addressing the diversity of ELL students by incorporating relevant variables such as native language and Limited English Proficient (LEP) status (Ardasheva, Tretter, & Kinny, 2012; Roberts and Bryant, 2011; Kim & Herman, 2009).

Calls for Research

Like most students, ELLs would experience more success in school if teachers effectively designed instruction to meet their learning needs. Based on the premise that ELLs are inherently heterogeneous individuals, various instructional and non-instructional elements could play an integral role in academic performance. Because Hispanic ELLs from lower socioeconomic backgrounds have received much of the attention in the literature, Genesee et al. (2005) made a call for research on different cultural and linguistic groups. They also indicated research is needed on middle school ELL students, especially those who are new to the country as adolescents. Research on older ELLs is so important because their learning curve is very different from ELLs who began their U.S. education in the primary years.

Similarly, Lesaux et al. (2007) suggested that conducting longitudinal research on ELLs through the middle and high school years would better inform whether previous gains and academic progress supported further achievement. Short and Fitzsimmons (2007) concurred the research on adolescent ELLs is sparse. In their report, *Double the work: Challenges and solutions to acquiring language and academic literacy for adolescent English language learners*, they reiterated the call that the diverse backgrounds of adolescent ELL students must be recognized because their cultural and educational differences should be considered when making instructional and assessment decisions.

The concern for recognizing the diversity of ELL students leads to collecting, maintaining, and using such extensive data in a comprehensive manner. When creating databases for K-12 students, multiple variables such as native language, educational history, and length of time receiving English instruction should be included (Wolf et al., 2010). Also, indicating that a student had participated in English as a Second Language (ESL) instruction would provide opportunities for comprehensive assessment of academic performance, especially that of long-term ELL students. Such background characteristics are particularly relevant to ELL students. Improved data management, such as longitudinal data collection, would be a start toward providing student-specific information as a basis for studying districts, schools, and classrooms as units of analyses. This design method would help produce stronger evaluations of academic programs and student achievement. LaCelle-Peterson & Rivera (1994) explained that progress over time could then be reported on grade cohort gains as well as in relation to specific variables with such data collection. These researchers further noted finding equity in assessment systems would depend on the accurate collection and reporting of student information, as well as the use of an assessment process that benefits students. Short and Fitzsimmons (2007) said many districts and states do not collect and analyze data on students who have exited ESL program services. They believe that “the true measure of a program or system's success is how well [ELL] students are doing in mainstream content classes" (p.16). Longitudinal studies using such informative data would offer valuable insight into which academic programs are working, and may also help determine what could be expected in literacy and English proficiency skill achievement.

Research Purposes

The intentions of this research are to emphasize the diversity of ELL students, use a hierarchical linear modeling approach to reveal that heterogeneity, and suggest methods for better data management. The first research purpose is to examine how initial LEP status impacts academic achievement. The second purpose is to explore which student background characteristics, including native language, country of birth, and number of years attending U.S. schools, affect reading and math performance.

Several variables were not used in the analyses, including ethnicity and sex, because past research had given them much consideration (Ardasheva et al., 2012; Kieffer, 2008; Robinson, 2008; Abedi & Lord, 2001; Wang & Goldschmidt, 1999). Furthermore, student-level socioeconomic status (SES) data were not available for use. This study refers to students whose first language is not English as English Language Learners or ELLs; these students are learning English language skills. The term Limited English Proficient, or LEP, is used here when discussing a student's status in the federally-reported subgroup. The LEP term will be discussed further in the Legislative Background section.

Research Questions

1. How does 4^{th} grade LEP Status relate to academic achievement in math and reading at initial status and over the following four years?
2. In what ways do native language, country of birth, and number of years attending U.S. schools affect math and reading achievement of elementary and middle school ELL students?

Research Hypotheses

1. H_1: LEP Status in 4^{th} grade will contribute to academic performance in math and reading at initial status and rate of growth.

2. H_1: Elementary and middle school ELL students will demonstrate diversity in how native language, country of birth, and number of years attending U.S. schools affect math and reading achievement.

These research questions and hypotheses address important substantive questions. In addition to learning more about ELL students, this study also makes a contribution by highlighting an important methodology and by raising awareness about the necessity of rich data structures for analysis.

ELL Populations

ELL students are diverse in their native languages, schooling background, and lived experiences. Although schools have progressed since the pre-NCLB days when linguistically-diverse students were relegated to the margins, clearly there is more work to be done to fully address the educational needs of ELL students. When this diversity is recognized more fully in the literature, the knowledge gained should help inform practice in educational settings.

Such diversity is present within North Carolina, the site of this research. Refugees and asylees have settled in numerous immigrant communities within North Carolina. With their resettlement beginning in 1986, the largest Montagnard community outside of Vietnam now resides in this state. The African immigrant/refugee population in one North Carolina county has grown to over 10,000 people. And, the central area of the state was cited as the third fastest growing Latino/Hispanic region in the United States, with over 75 percent being Mexican immigrants and their families (Saavedra, Morrison, Smith, & Bailey, 2008). Most of these children began their U.S. schooling as English language learners, whether they arrived from another country or lived in a home where English was not the first language. As a result of these students' linguistic diversity, the ability to conduct this research is tied to the availability of such rich data and heterogeneous subsamples.

Methodological Approach

Examining the diversity of ELL students, by incorporating contextual variables with limited sample sizes, can be accomplished by using hierarchical linear modeling (HLM). Such an integrated approach can highlight the variability that exists between students which, in turn, can help provide more nuanced understandings of individual predictors. Formulation of models for individual change in this hierarchical approach can be an important step toward later recognizing how external influences, such as instructional methods and school environments, affect academic achievement over time (Raudenbush & Bryk, 2002).

This study considered background characteristics that are relevant to ELL students, including native language, country of birth, and number of years attending U.S. schools. To date, few studies have included these specific variables in rigorous analyses and studying them would capture a purer picture of student achievement. Another important component of this study is that the data feature four years of math and reading scores for a cohort of English language learners from the same school district, following students from 4th grade in elementary school to 7th grade in middle school. Such a longitudinal approach using complete data ensures a consistent sample with stable background variables. These repeated observations are not independent, but HLM can handle this nesting well and allow for the study of within-student and between-student variation in math and reading achievement. By incorporating student-specific variables with an appropriate analysis like HLM, a more complete picture of ELL students' academic performance over time can be achieved.

Database Management

While the current results cannot be compared across states due to the state-specific test data used, these findings are framed to demonstrate the value of managing and using relevant student background information to study the academic achievement of ELLs over the course of their educational careers. Learning more about students who have exited ESL instruction could be just as informative as learning about students currently receiving services (Lesaux, 2006). Longitudinal within- and between-state studies could also be conducted as suggested by Short & Fitzsimmons (2007). States that have newly implemented a common core curriculum and administer common assessments could consider developing such a research platform by sharing and consolidating student data.

Ultimately, heterogeneity will be recognized and a more complete picture of ELL students' academic achievement will be presented. The subsequent awareness and knowledge could aid educators in creating and implementing educational programs that could specifically meet the individual needs of ELLs. This research intends to broaden the literature on the academic performance of ELL students, in the context of appreciating their diversity, by using a hierarchical linear modeling approach featuring relevant and meaningful variables.

Legislative Background

The reauthorization of the Elementary and Secondary Education Act (ESEA) by the No Child Left Behind (NCLB) Act in 2001 required states to administer yearly academic assessments and increased accountability for student subgroups. In 2012, the U.S. Department of Education approved North Carolina's request for a flexibility waiver from some of those requirements so that additional focus could be placed on increasing the quality of instruction and improving student learning. As a result, states developed annual measurable objectives (AMOs) which are performance targets that states, districts, and specific student subgroups must achieve each year to meet the NCLB requirements. LEP is one of those federally reported subgroups, in which ELL students are included based on their level of English language proficiency.

In North Carolina, students are identified as LEP based on results of the WIDA-ACCESS Placement Test (W-APT), an English language proficiency screener test designed by World-Class Instructional Design and Assessment (WIDA). WIDA is a non-profit group that develops standards, instructional resources, and assessments to promote educational equity for English language learners. North Carolina has been a member state of the WIDA Consortium since 2008 and has adopted English Language Development Standards and annual assessments for reporting purposes.

Students who are classified as LEP participate in ESL instruction in North Carolina, which may be delivered in the form of pull-out activity, separate class, or in-class assistance. Students exit ESL services when they score accordingly on WIDA's ACCESS for ELLs, an English language proficiency assessment. ACCESS for ELLs is not intended as sole justification in determining a student's language proficiency; however, no other criteria, including performance in academic content, are considered during the reclassification process in North Carolina. As Chalhoub-Deville & Deville (2008) clearly pointed out, “Because students sound proficient in everyday communication does not mean they are proficient in using the language to perform more cognitively demanding and abstract tasks, i.e. to use appropriate language in the school and classroom environment” (pg. 514). English is still not the ELL student's first language, and never will be.

English language learner students of all ages arrive in the U.S. and their academic experiences and needs differ. Based on the categorization from Olsen & Jaramillo, Freeman & Freeman (2002) describe three types of older English learners who attend upper-elementary, middle, and high schools. "Newly arrived with adequate schooling" students have been in U.S. schools for less than five years but received sufficient schooling in their native country and can usually catch up academically. Students who are considered "newly arrived with limited formal schooling" have also been in U.S. schools for less than five years, but their prior schooling has been interrupted or limited and they have poor academic achievement.

The third group of older English language learners includes students who are considered long-term English learners (LTELs). These students have attended U.S. schools for seven years or more, are below grade level in reading and writing, and score low on tests. In contrast, many long-term English learners receive passing grades on their work and have conversational fluency in English, which leads to a false sense of academic achievement (Freeman & Freeman, 2002).

The Californians Together coalition conducted a statewide survey on high school English learners in 2009, which resulted in the report *Reparable Harm* (Olsen, 2010). Of the 40 school districts that participated in the survey, 59 percent of their ELL high school students were considered long-term English learners. Several factors were found to contribute to this situation during their schooling experience, including social segregation, linguistic isolation, and instructional materials that were not designed to meet their English learner needs. The California Assembly passed a bill in 2012 (AB 2193, Lara, Long-term English learners) which provided a clear definition in Section 1. Section 313.1: "'Long-term English learner' means an English learner who is enrolled in any of grades 6-12, inclusive, has been enrolled in schools in the United States for more than six years, has remained at the same English language proficiency level for two or more consecutive years as determined by the English language development test…" This report and legislation are important steps toward recognizing and, eventually, addressing the special needs of older English language learners.

For accountability purposes, students who have exited LEP identification are still considered members of the LEP subgroup for two years after leaving ESL services. But, after those two years, their test score data are no longer linked to the LEP subgroup. Subsequent academic achievement, still as a student with English as their second language, is therefore difficult to consistently track and assess.

This inability to follow ELL student performance is notably problematic because the LEP classification is not consistently defined across or even within states (Abedi, 2004; Abedi, 2008; Goh, 2004). States and school districts use different methods to identify students for inclusion and to exit students from the subgroup (Wolf et al., 2010). Without a common definition for LEP classification in place, it has not only been difficult to make comparisons and propose generalizations but also to universally track the academic performance of ELL students over time.

Theoretical Basis

Delpit (2006) and Deyhle (2009) both compellingly discuss how the dominant group views a person of one race as a representative of the entire race. Similarly, in schools, some educators choose to wrap up all ELL students into the same package, and consider them an “undifferentiated mass” (Delpit, 2006). Along the same lines, educators go so far as to overtly assume and stereotype that if the student appears to be Hispanic, then the student must be an ELL as well. If educators consider some students as a solitary group based on their perceptions of students' cultural identity or the students' native language, then students' individual needs cannot be equitably met. Bennett (2006) supports this view when she wrote, “Potentials may differ, and at times equity requires different treatment according to relevant differences” (p. 18).

By not seeing each ELL as an individual student, educators fail to recognize and give value to each student's culture and identity. For example, by emphasizing English and trivializing the Spanish language, educators give the impression that the student's native language is irrelevant or substandard. Moreover, educators often view ELLs as being inferior and; therefore, treat them as others (San Miguel & Donato, 2009; Schoorman, 2001). This behavior insinuates something is wrong with the student's family and culture (Delpit, 2006; Purcell-Gates, 1995). The language a student brings to school is an integral part of who that student is. Language is a connection to families and culture (Delpit, 2006: McBrien, 2005). Macedo (2006) pointed out that, "…language may either confirm or deny the life histories and experiences of the people who use it" (p. 131). Denying a student's native language is much like denying their value as a person and as a member of their family and culture. A student's native language is the means through which they develop their own voice and make sense of the world (Macedo, 2006).

When the research can fully recognize ELL students' diversity, then that diversity can become an asset for learning. The school environment would be such that multiple languages and cultures are embraced and valued as resources (Cummins, 2009), rather than being treated as signs of deficiency or otherness. WIDA's first two *Guiding Principles of Language Development* (WIDA, 2010) aptly support this sociocultural approach:

1. Students' languages and cultures are valuable resources to be tapped and incorporated into schooling.

2. Students' home, school, and community experiences influence their language development.

By using rigorous analyses and relevant data, this research is intended to increase awareness about the varied educational and social needs of linguistically- and culturally-diverse students.

CHAPTER II

LITERATURE REVIEW

This chapter presents two sections of existing literature that are relevant to the research problem. The first section is a discussion of studies in which non-hierarchical linear modeling analyses were used to examine the academic achievement of ELL students. The second section focuses on studies where researchers did use a hierarchical linear modeling approach. Both sections describe the relationship of one or multiple background variables to academic achievement of ELL students. Findings from research that compared the academic performance of current and exited LEP students will also be discussed. The acronyms ELL and LEP are often used interchangeably in the literature. In this review, though, ELL will refer to a student whose first language is not English and non-ELL to a student who is a native English speaker. Also, LEP will refer to an ELL student who qualifies for and attends ESL instruction. And exited, former, or non-LEP will refer to an ELL student who no longer qualifies for nor attends ESL instruction. If the researcher specifically named their sample groups, such as former-ELL, then those definitions are described and used in the respective summaries. Additional acronyms and definitions are presented in Appendix A.

Non-HLM Designs

While English Language Learners share learning English as a commonality, these students are inherently different with their varied backgrounds and experiences (Goh, 2004; Chalhoub-Deville & Deville, 2008). Their differences include cultural background, ethnicity, educational experiences, family history, socioeconomic status, and native language. Some research studies have demonstrated differences in academic performance for ELL groups with varied cultural and language backgrounds. Traditionally, though, most studies focused on the academic performance of current LEP students, of which comparisons were often made with native English speakers by using a broad set of methodologies.

Abedi and Lord (2001) used National Assessment of Educational Progress (NAEP) math items to compare the academic achievement of non-ELL and ELL students. They also considered whether background variables, including sex and family SES, impacted math performance. The dataset featured 1,174 8th grade students from 39 classes in 11 Los Angeles schools, of which 802 were non-ELLs and 372 were ELLs. The researchers identified the sample's ELL students' language classifications which included various levels of English proficiency (Initially Fluent in English, Redesignated Fluent, Limited English Proficient, other categories of English as a Second Language). To test for a difference in math performance between ELLs and non-ELLs at low and high SES levels, a two-factor ANOVA model was applied to the data. Abedi and Lord found that the ELL students' mean math score was significantly lower than the mean score for non-ELL students. Also, low SES students performed significantly lower than those in the high SES group.

Native language data were made available from Language Background Questionnaire and the languages included Spanish, Korean, Chinese, Farsi, and Filipino; however, the native language variable was not used and counts for each language were not stated. The failure of the researchers to not use the native language data may have limited their ability to draw specific conclusions about ELL students.

Callahan (2005) used linear regression models to identify significant predictors and the amount of variance in a specific academic outcome for each of those predictors. The study's purpose was to determine if track placement or English language proficiency predicted academic achievement of high school ELL students. Academic achievement was explained by access, or lack thereof, to content. One independent variable of interest was recent immigrant status, which was defined as being enrolled in U.S. schools for five years or less. The data were collected from 355 English language learners in one rural high school in California. Eighty-nine percent were Spanish speakers and the remaining students spoke one or more of 11 other languages. After Spanish, the next two populous languages were Punjabi and Urdu. Although native language was described, its relationship to academic achievement was not examined in the research. The sample was divided into three groups: long-term English language learners, recent immigrants with a large amount of previous schooling, and recent immigrants with a limited amount of previous schooling.

Callahan found that recent immigrant status was significant in predicting GPA, credits, and SAT9 Math, but was not significant for SAT9 Reading, California High School Exit Exam (CAHSEE) language arts, or CAHSEE math. Also, English learner cohort was significant in predicting academic achievement in both content-area and language-based measures. Long-term, as opposed to recent immigrant status, proved to be significant in predicting performance in terms of grades, credits, and SAT9 math. English proficiency level was significant only in predicting achievement on SAT9 reading and CAHSEE language arts. Finally, recent immigrants with a large amount of previous schooling enrolled in a slightly higher proportion of college-preparatory coursework than did other ELL students.

Bankston & Zhou (1995) surveyed nearly 400 Vietnamese students in two New Orleans high schools to learn how student's literacy in their native language contributes to academic achievement. The researchers selected this particular group for study because they believed the students could offer insight into the educational experiences of a new immigrant population in a disadvantaged minority environment. They found strong correlations between ethnic identification and native language literacy, as well as relationships between academic effort and both native language literacy and self-identification. Those students who reported reading and writing well in Vietnamese were much more likely to indicate receiving high letter grades than those students who did not report having good reading and writing skills. Based on their findings, the researchers suggested that literacy in a native language should be encouraged, rather than tolerated, as those linguistic abilities contribute to overall academic achievement.

In a statistical profile, Stiefel, Schwartz, & Conger (2003) reported that standardized test scores varied significantly among ELLs across different language backgrounds in New York City. They listed various background characteristics by language groups, including percentages of Limited English Proficient, poor (as determined by free or reduced lunch eligibility), immigrant, ethnicity, and special education. The home language groups featured were English, the top ten non-English languages, and "Other." This final category, "Other," was assumed to contain all the students' native languages outside the top ten, but that definition was not provided in the report.

By using z-scores from standardized test administrations, Stiefel et al. found that Chinese-Dialect (1.036) and Korean (1.028) speakers scored more than one standard deviation above average in math, while Haitian-Creole (-0.306) and Spanish (-0.252) speaking students scored well below average. These results show that academic achievement may be linked to native language or ethnicity; however, similar studies have not been completed with more comprehensive longitudinal analyses.

A study by Hopstock & Stephenson (2003) was funded by the US Department of Education to provide more detailed descriptions about differences among districts, schools, and LEP students based on native language group. Collected from 932 questionnaires, the data were weighted to be nationally representative of districts and schools serving LEP students. The analyses were reported for three groups: Spanish as the majority, another language group, or no language group as the majority. Hopstock & Stephenson defined a majority language as one which represented at least 50 percent of all LEP students in the district or school. The researchers stated the number of students speaking each language and the number of different native languages in districts. They indicated that they had considered analyses of other specific language groups or combinations of languages; however, it was decided the language groups were not large enough to do so. Their explanation was that because Spanish represented a large majority (76.9% or nearly 3 million students) and no other language was more than 3 percent of the sample, that it was best to create two native language groups: Spanish and other. The "other" category included more than 34 languages. This study was purely descriptive in that percentages were noted to make comparisons about districts' instruction and services. Because of the report's intent to be descriptive, it would have been more informative to provide detail about multiple native language groups, whether individually or in meaningful language groups.

Stevens (1999) accessed the 1% Public Use Sample of the 1990 U.S. Census to show relationships between social characteristics and English proficiency levels among adult immigrants. The census data did not include native language information. Stevens inferred the immigrants' prior knowledge of English based on whether English was a dominant or official language of their country of birth, and if not the case, then those immigrants' data were included in the sample. Although more current data would be available today, it would be impractical to make similar native language assumptions because many ELLs are born in the United States, including 62 percent of the Spanish speakers in this dissertation sample.

Ordinal logit models were used to analyze the data so that the ordering of the census categories would be taken into account. Several variables, including age at immigration and born in a Spanish-language country, were included in one model to predict level of English proficiency. The age at immigration variable was estimated by using the time period of immigration and the respondent's age at the time of the census. The country of birth variable was described in binary terms, whether or not born in a Spanish-speaking country. If not born in a Spanish-speaking country, the term was deemed "non-anglophone country."

Stevens founds that immigrants from Spanish-language countries reported lower levels of English proficiency than immigrants from non-anglophone countries. She also found that proficiency in a second language was strongly affected by age at immigration. Respondents who began to learn English prior to the age of five self-reported being English proficient as adults. However, Stevens reasoned that part of the success could be tied to other demographic and social characteristics that may be related to arriving in the United States at an early age.

Abedi, Lord, Hofstetter, & Baker (2000) looked at ELL student performance on math word problems, while considering the effect of accommodation strategies and the impact of various background characteristics on the level of effectiveness of different accommodations. Some of these student background variables included type of math class, form of accommodation, country of origin, spoken language other than English, television viewing habits, attitudes toward math, and language of instruction. The research team administered NAEP items along with four different accommodations: modified items (simplified English), glossary, extra time, and glossary with extra time.

In the sample of 946 8th graders, the "county of origin" variable was grouped by United States (57.1%), Mexico, and Other countries. Additional details on the "Other" countries were not provided. LEP status was classified into three groups: Limited English proficient (LEP), Fluent English proficient (FEP), and Initially fluent in English (IFE). For the analyses, Abedi et al. kept LEP (52.8%) as one group and considered these students ELL. However, the FEP (30.4%) and IFE (16.8%) groups were combined and defined as non-ELL because the FEP students had transitioned into non-LEP programs and the IFE students were initially fluent in English. For the "speak other language" variable, 85.1 percent of the sample noted they spoke one language in addition to English and 82 percent indicated Spanish. So, individual student native language data were apparently available, but not used. The number of years the students have lived in the U.S. was stated in ranges (1-3, 4-6, 7-9, 10-12, and more than 12 years), and again, although the information was available, it was not of interest in this study.

Abedi et al. examined whether specific accommodations helped some students more than others. One multiple regression model included all independent variables and a restricted model used only variables representing main effects. They discovered that the full model had more predictive power and explained a larger amount of the variance than the restricted model. A discussion on three predictors that were significant at the .01 level was presented; however, the article failed to note and detail a fourth predictor, country of origin, as a main effect (*p* = .014). This is an important omission considering that country of origin is generally a unique characteristic to English Language learners.

Math performance across the accommodation categories differed significantly; students also performed differently across the two ELL status categories. The researchers reported that the trends remained stable after controlling for reading achievement scores. They found that the combined effects of accommodation strategies and background characteristics were more powerful predictors of student performance than the strategies and characteristics examined separately.

deJong (2004) conducted a regression analysis to determine if length of bilingual/ESL program attendance and program exit grade level had a significant effect on achievement outcomes in reading, math, and science of "former ELL" students who had attended either a bilingual or ESL program. A former ELL was defined as a student first classified as limited English proficient upon school entry, attended a bilingual or ESL program, then reclassified as fluent English proficient and enrolled fulltime in a mainstream classroom. The researcher intentionally chose to use the term "former ELL" because their academic performance was no longer tracked as a former ESL program participant and they were perceived by others as native English speakers.

The sample size was limited, just 38 4th graders and 56 8th graders, and the researcher acknowledged the limitations using a small sample size. After controlling for program model, exit grade level was found to be a significant predictor for 4th grade ELA and Science, in that the higher the grade level the student exited the program, the lower their 4th grade scores. For the exited ESL program students, the lower the grade level that they exited the program, the better their 8th grade science scores. Conversely, exited bilingual program students' science scores slightly improved if they exited at higher grade levels.

These research studies have contributed to the literature in valuable ways, such as incorporating variables specific to ELL students and taking a longitudinal approach, but there is so much more to learn about these students' varied cultural and linguistic backgrounds. Although attempts were made to recognize students' native language, the focus was generally on Spanish speakers. In some instances, individual student data were available but not used in the analyses, or research was limited by sample size.

HLM Designs

A more meaningful and informative picture of ELL academic performance can be realized by examining test score observations over time within a hierarchical linear modeling design. The following research studies incorporated this approach, and most designs delved specifically within ELL student achievement.

Lesaux, Rupp, & Siegel (2007) followed the reading development of 824 students in a Canadian school district from Kindergarten through fourth grade. Their research purposes were to investigate differences in reading achievement of ELLs and non-ELL peers at kindergarten and Grade 4, examine kindergarten predictors of ELLs' and non-ELLs' fourth-grade word reading and reading comprehension, and model developmental trajectories of word reading for the two groups. Multiple analysis of variance, hierarchical linear regression, and growth modeling were used to address the three research questions. In the first model, the mean proficiency level was modeled without predictors and varied across schools. Language status was added as a predictor at Level 1 in the second model to detect differences between the two groups.

The sample included 689 native English-speaking, and 135 English language learner students representing 33 different languages, of which Cantonese, Mandarin, Korean, Spanish, Polish, and Farsi were the predominant native languages. Although the researchers had native language data, they did not use it in their analysis, perhaps due to the small ELL group size and the numerous native languages.

Lesaux et al. found only slight differences in the reading development of ELLs and non-ELLs. In kindergarten, the ELLs performed more poorly than the non-ELL group on several tasks of early literacy, but by fourth grade, those differences had generally disappeared. The ELL students performed similarly to, and in some cases, better than, non-ELL students on all but one fourth-grade task. The nonlinear developmental trend in word reading from kindergarten through fourth grade was very similar for both groups. Another finding for both groups was that letter identification skills in kindergarten were predictive of initial mean differences and growth over time in word reading.

Lee & Madyun (2008) explored the relationship between racial composition and academic achievement by combining two groups, Hmong and LEP students, who have often been marginalized in both research literature and society. School district results from the 2002 Metropolitan Achievement Test were used in a two-level HLM design, which included standardized reading scores for all Hmong students (n=1622) enrolled in

St. Paul, Minnesota public middle schools. Of those students, 83.8 percent were receiving ESL services and classified as LEP.

Results indicated that the LEP variable was the most important predictor in determining Hmong student achievement. When all student-level variables (sex, special education status, SES, and LEP) were controlled, non-LEP Hmong students tended to earn higher reading scores than LEP Hmong students. A school-level predictive variable, different race exposure, was included in the Level 2 model. The student-level predictors remained statistically significant and, in addition, a cross-level interaction occurred between LEP and different race exposure which indicated a positive association between Hmong LEP student achievement and different race exposure. On the other hand, reading achievement was negatively associated with same race exposure. The more diverse a school became, the higher the academic achievement of Hmong LEP students.

Kieffer's (2008) study on reading development trajectories of ELL students and native English speakers used data from the Early Childhood Longitudinal Study, Kindergarten (ECLS-K) dataset. Three sample groups were created using the students' initial classification from parental report and English language assessment scores: ELL students who entered kindergarten with limited English proficiency, ELL students who entered kindergarten with full English proficiency, and native English speaking kindergartners. Of particular note in this study, the native language variable was constructed as binary (Spanish speaking/not-Spanish speaking) with 63.6 percent Spanish speakers in the cohort. No explanation was provided to indicate the reason for excluding other languages.

The student-level variable SES and all school-level variables (SES, students of color, and LEP) were considered time invariant and fixed to the average value across the five observations, while ethnicity and native language were not fixed. A quadratic growth specification was used to represent the individual growth trajectories. Random effects were included in the model and represented Level 2 residuals for the intercept and slope. Due to convergence problems, a random effect for the acceleration term could not be included and it became fixed across individuals.

Kieffer found that the students with limited English proficiency had growth trajectories that presented at much lower elevations than did the native English speakers throughout their elementary years, with greater differences in achievement by the 5th grade. The fully English proficient students had trajectories that were similar to the native English speakers. Furthermore, controlling for SES and other demographic factors reduced the effect of initial English proficiency and yielded differences that narrowed over time.

D'Angiulli, Siegel, & Maggi (2004) examined socioeconomic gradients and growth-mixture trajectories of word-reading achievement for 1,108 students in one school district who received literacy-intensive instruction beginning in kindergarten. These gradients referred to each grade level's mean reading scores regressed onto the SES indicator. Reading trajectories were obtained for each of the four SES quartiles which allowed for investigation of differences in developmental patterns within the same SES level. They found that, in kindergarten, the relationship between SES and word reading was significant in two of the three subgradients identified in English language learners, and in the only gradient identified in native English speakers. In the lowest and highest SES groups, ELL students had significantly lower word-reading scores than native speakers. Yet, by grade 3 or 5, the ELL students improved more than the native speakers, even though in kindergarten they were the most at risk for reading failure. The researchers suggested that the literacy-intensive program may have reduced the negative influence of SES on word-reading development. Interestingly, ELL and non-ELL students at the middle-SES level improved similarly as they progressed through grade 5.

Roberts, Mohammed, & Vaughn (2010) were interested in finding strong estimates of reading achievement patterns between ELLs and native English speakers but also within ELL students using data from the Early Childhood Longitudinal Study (ECLS) sample. The three groups of interest were native English speakers, Spanish-speaking ELLs, and Asian-language ELLs. Only students who were proficient in oral English at the end of kindergarten were included in the study. Because detailed information was only provided for Spanish and English student languages in the public version of the ECLS data, the researchers chose to make assumptions about Asian-language speakers in their sample. As a result, Asian race/ethnicity students classified as ELL were coded with an Asian-language status, and those students formed the third group.

To examine the main effect of native language and its relationship with SES, means and variances were compared from difference testing of nested models. A single-group hierarchical model was fit to isolate the effect of native language on early reading achievement and estimate the effects of SES and other school-level variables on reading achievement.

This study by Roberts et al. indicated that achievement trends of Asian-language ELLs were more similar to those of native English speakers rather than Spanish-speaking ELLs. Spanish-speaking ELLs had lower initial reading achievement than both Asian-language ELLs and native English speakers, while Asian students had higher initial achievement than did the native English speaking group. Additionally, Spanish ELLs demonstrated statistically significantly less growth over time than did Asian-language ELLs, with differences being most notable on reading evaluation-related tasks. However, language-related differences in overall reading were reduced when SES effects were specifically modeled, which suggests that SES may be a major factor in explaining the lower achievement rates of English-proficient native Spanish speakers. Spanish-speaking students were able to keep pace or quickly catch up in word-level skills areas. However, they lost ground to native English speakers and to Asian-language students over time on text-level skills, which resulted in increasingly large comprehension deficits.

Also using ECLS data, Roberts and Bryant (2011) estimated math achievement trends of students from kindergarten through 5th grade by evaluating the effect of low SES and native language group specific math skills. The student groups of interest included native English speakers and ELL students who were English proficient by the end of kindergarten. The ELL groups were categorized as either native Spanish speakers or speakers of Asian languages. In this dataset, primary language was not specified for students speaking a language other than English or Spanish at home. Therefore, the researchers assumed Asian-language status for all students coded as ELLs and Asian in the race/ethnicity fields.

The results of the latent variable growth modeling showed that native language may be less relevant than SES for predicting math achievement. Math-related school readiness was correlated with SES across various math skills for all three student groups. When assessing the effect of primary language on future math achievement, the groups differed in their levels of readiness with the Spanish-speaking ELL group scoring lower than both the native English-speaking and Asian-language ELL students.

Using ECLS-K data and longitudinal analysis, Chang (2008) observed the effects of teacher classroom grouping practices on math achievement of language minority students. The study examined the performance of four ethnic groups (Caucasian, African-American, Hispanic, and Asian) and three language groups: English only, dual-language (with no difficulty speaking English), and ELL (pull-out or in-class ESL program), and is one of few studies to specifically address the heterogeneity of ELL students in a longitudinal research design. Separate analyses by ethnic group were conducted in order to avoid variable collinearity and to provide clarity for interpretation. Specific details on whether dual-language students ever participated in ESL instruction were not stated, and could have provided additional insight in the analyses.

The baseline model was used to compare the growth trajectories of math performance. Results showed that the ELL group displayed lower math performance than did English-only students in the Hispanic and Asian groups. Caucasian ELL students demonstrated significantly lower performance than Caucasian English-only students during the Kindergarten spring semester. However, by grade 5, the Caucasian ELL students narrowed the performance gap that separated them from English-only Caucasian students, demonstrating a significant growth rate. The Hispanic ELL group began with a significantly lower math score and increased their math scores at a significant slower pace, compared with the English-only group.

Chang's math grouping model examined the differential effects of four instructional grouping practices on student performance. In teacher-directed whole-class activity, Caucasian and African-American English-only students displayed increased math achievement scores, while Hispanic ELL students experienced a negative effect. Teacher-directed small-group activity was found to positively influence Caucasian dual-language students but negatively influence Asian ELL students. Teacher-directed individual activity showed a significant positive result for math performance of the Hispanic dual-language group when compared to the Hispanic English-only students.

Robinson (2008) used ECLS-K data to examine whether ELL students would benefit from ability grouping for reading instruction more than children from English-speaking home environments. This research used student- and school-random effects in a hierarchical linear model and difference-in-differences estimators to determine whether an overall effect of ability grouping occurred in kindergarten and first grade and if a differential effect existed for some race-language groups. This model allowed each in-school period, such as kindergarten and first grade, and the summer period between kindergarten and first grade to have its own growth differential.

Seven race-language groups were defined. Three of the race-language groups were not different in terms of primary home language: White, non-Hispanic; Black, non-Hispanic; and other, non-Hispanic, which included Native Alaskans, American Indians, Native Hawaiians, and other Pacific Islanders. A language minority/primary (LM/P) student was defined as a student from a home where a language other than English is routinely used and that language was the one primarily spoken. The four remaining race-language groups were separated Hispanics and Asians on the basis of primary home language: Hispanic, LM/P (n = 238); Hispanic, non-LM/P (n = 588); Asian, LM/P (n = 162); and Asian, non-LM/P (n = 159). The researcher explained that there were too many different Asian languages which resulted in very small sample sizes, and so the research focused on the differences between LM/P and non-LM/P Hispanic students. It was unclear whether the Hispanic students all spoke Spanish and that was why they remained in the analyses, compared to the multiple Asian languages spoken by the Asian LM/P group.

LM/P Hispanic children were found to likely benefit from ability grouping in kindergarten and first grade and that ability grouping could help greatly reduce the achievement gap that they face upon entering kindergarten. Yet, gains made by ability-grouped LM/P Hispanic students in kindergarten tended to weaken over the next year unless they were also grouped in first grade. If they continued to be ability-grouped, the benefits of grouping then accumulated.

While respecting the individual Asian languages would have been a preferable method of analysis, the study may have missed an opportunity to explore the heterogeneity of ELL populations by not using the Asian LM/P student data as one group and comparing those results to the Hispanic LM/P group.

Sung & Chang (2008) used 4 waves of IRT scale scores from the ECLS-K dataset in a longitudinal hierarchical comparison of the differential effects of care programs on reading achievement growth of three groups of students: native English speakers, non-LEP, and LEP-at-K. The non-LEP students were those who spoke a non-English language at home but were not placed in ESL classes. Those students who attended in-class or pull-out ESL instruction during kindergarten were grouped as LEP-at-K. It was found that language minority students in the USA displayed comparatively high reading performance when they had continued participation in center- or school-based care programs during the early years of schooling, while native English-speaking students did not display those patterns from the same conditions.

When comparing the reading performance of non-LEP (dual language) students to native English-speaking students, Sung and Chang found that the non-LEP students began school with significantly lower scores than the native English speakers. However, the non-LEP students had a higher initial growth rate which indicated they were catching up with their English-speaking peers. In terms of participation in a center-based program, significant positive reading effects were found for both the non-LEP and LEP-at-K groups; however, reading scores for the English-speaking groups were not observed.

Hofstetter (2003) used hierarchical linear modeling to examine how contextual factors, particularly at the classroom level, influence LEP and non-LEP 8^{th} grade Latino students' performance on the NAEP mathematics assessment. The LEP group (n = 676) was defined based on their English language proficiency scores; the non-LEP group (n= 173) included students who were initially fluent or currently considered fluent English proficient. The sample featured students with a variety of native languages, such as Spanish, Cambodian, Khmer, Vietnamese, Hmong, Lao, Thai, and Farsi. However, only students who self-reported themselves as Hispanic, who identified Spanish as a second language, who came from a Spanish-speaking country, or who claimed both of the last two characteristics were included in the sample.

One initial variable of interest was Years in United States, which ranged from less than 1 year to 14 years; however, no significant predictive value was found and it was deleted from the model. Four other student-level variables were also deleted, one of which was the number of times the student changed schools in the last two years.

In Hofstetter's study, students were administered one of three different math test booklets: modified English accommodation, original Spanish accommodation, or original English standard edition. The LEP students who received math instruction in Spanish performed lower on the English standard edition than did students instructed in English, once key variables were held constant. On the other hand, LEP students instructed in Spanish who took the Original Spanish accommodation test performed higher than students with no accommodation.

Slama (2012) used growth modeling to fit a hierarchical linear model in order to examine the heterogeneity in academic language development over students' high school trajectories on the basis of ELL generational status. The term "generational status" was defined as whether a student was U.S. - or foreign-born. The article also stated that it was assumed foreign-born students had spent less time in U.S. schools than U.S.-born students. Data on actual length of time spent in U.S. schools were not available for use, which was acknowledged as a study limitation.

Five waves of 9th through 12th grade academic English proficiency data from a statewide cohort of ELLs (*n*=3,702) were used; students were nested in 193 schools and the random effects of school were included in the intercept of all the models. In the sample, 58 percent of the high schools students (*n* = 2,144) were Spanish-speaking and 59 percent were U.S. born. Although the research focused on the academic English proficiency of U.S. - versus foreign-born ELLs, individual native language data were available for use.

Slama's results indicated that U.S.-born ELL students exhibited higher levels of academic English proficiency at the beginning of ninth grade than did their foreign-born ELL peers. However, foreign-born ELL students exhibited faster rates of growth than did U.S.-born ELLs and, by the end of high school, had caught up to their native-born peers on measures of academic English proficiency.

Maerten-Rivera, Myers, Lee, & Penfield (2010) examined both student and school predictors of science achievement from a culturally and linguistically diverse sample of nearly 24,000 5th grade students from 198 schools in a large urban school district. They hypothesized that student-level factors, such as sex, ethnicity, socioeconomic status, and ESOL status, would impact student science achievement at the elementary level. All of the student-level factors were included in the in model to examine the unique effect of each, while also controlling for the effects of the other variables.

Two groups were considered in the ESOL variable. For the first group, students at ESOL Levels 1-4 who were currently in ESOL programs were coded as *ESOL*. The second group included students at ESOL Level 5 who had exited ESOL programs within the past two years but were being monitored for two years, those who had exited ESOL programs more than two years ago, and those who were never in ESOL. With this group design, the performance of never-ELL students was considered in the same group as students whose first language was not English.

Maerten-Rivera et al. reported that the effects of sex, ethnicity, and SES on science achievement were small. However, ESOL status had the largest effect, in that, on average, a decrease of about 23 points in science achievement for an ESOL student would be expected compared to a non-ESOL student. This result was able to be determined because the researchers chose to include all relevant factors in their model in order to find the unique effect of each while controlling for the effect of the others. The relationship between reading and mathematics achievement with science achievement was also explored. The interaction term between ESOL and reading achievement was a significant predictor of science achievement, but the interaction term between ESOL and mathematics achievement was not significant.

Quantile regression was the analysis of choice for Chen (2010) to explore the relationship between math achievement and language proficiency. Quantile regression can offer more information about the relationship between variables when observation of various locations on the entire distribution is preferred over looking at average performance (Chen & Chalhoub-Deville, 2014). Demographic information from the ECLS-K dataset was used to create the three ELL sample groups: native English speakers, former ELLs and ELLs until spring 1^{st} grade, and ELLs after spring 1^{st} grade. Assessment data from grades K, 1, 3, 5, and 8 were used in the research.

Chen's analysis indicated that reading scores explained the gap in math achievement between English Language Learners and native English speakers. She also found that language influence affected math achievement differently between students with different math ability, in that the relationship was stronger for students with low math ability and weaker for those with high ability. However, the strength of the relationship decreased as students moved to higher grade levels.

Wang & Goldschmidt (1999) used a growth model to examine students' language proficiency and math achievement. They found that 8^{th} grade LEP students performed significantly worse in math than students who were native English speakers. Interestingly, another finding revealed that students who had been redesignated as English proficient after exiting the ESL program had a significantly faster math growth rate than did students who were native English speakers. On the other hand, redesignated students did not benefit as much from participation in honors courses compared to English-only students. The researchers argue that having sufficient English skills [to be classified as proficient] are not enough to be successful in advanced math courses.

Wang & Goldschmidt's dataset included three years of test scores and immigrant status, race, sex, and language characteristics for nearly 2,500 middle school students.

Language proficiency was categorized by three typical groups whereby scores on the California Test of Basic Skills determined designations: LEP (limited English proficient); Redesignated (former LEP students who had exited an LEP program); and FEP (fully English proficient).

Ardasheva et al. (2012) investigated how academic achievement in reading and math of middle school former ELLs compared to current ELL and native English-speaking students. They estimated a two-level model using maximum-likelihood estimation, and included student-level predictors such as sex, age, ELL status, and SES. The large sample featured 17,470 native English-speaking students, 558 current ELLs, and 500 former ELLs. A former ELL was defined as a student who was redesignated as fluent English proficient. The data from the three grade levels were combined to form a single middle grades sample. The aggregation was done because some schools lacked either current or former ELLs at some grade levels, which would have resulted in the exclusion of these schools from the analyses. Native language data were available in that current and former ELLs spoke 48 different languages; yet, the data were not used in the study and their use may have added more information to the findings.

After controlling for gender, age, and student- and school-level SES, results indicated that former ELLs significantly outperformed current ELL students and native English speakers on state reading and mathematics tests. And, former ELLs did not lose this advantage in higher poverty schools. Former ELL students in higher poverty schools did slightly better than did their former ELL peers in lower poverty schools.

Kim & Herman (2009) designed a three-state study with the intent of providing more accurate estimates of achievement gaps between ELL students and non-ELL peers by refining ELL student subgroups. Students were divided into four groups: current ELLs, recently reclassified students (reclassified as fluent in English within the last two years), former ELLs who were reclassified more than two years earlier, and non-ELL students. By separating the former ELLs from the non-ELLs, they hoped for better estimate differences in achievement between current ELLs and other students, while also learning more about achievement differences within the ELL population in the process. One limitation was that each state used different methods to evaluate and reclassify ELL students.

Using data from three different states produced a large sample size which included over 150,000 students just for the grade 8 analyses. Two-level hierarchical models were fit to estimate average achievement gaps for scores for each combination of states, grades, and content areas. However, each state used different assessments. So, to make within- and between-state comparisons possible, state-by-state analyses were conducted and results were converted into standard deviation units which then allowed for comparisons between content areas and grades.

Kim & Herman found statistically significant achievement gaps between current ELL students and their non-ELL peers in all three states, although the magnitude of the gap varied by content area. For example, a smaller gap occurred in math while a larger gap was seen in reading and science. Mixed results were seen for recently reclassified students, whereas in one state, those students performed lower overall than non-ELLs but in another state, their performance was better than non-ELLs. In terms of grade level achievement within states, the middle school current ELLs, recently reclassified ELLs, and former ELLs performed worse than their elementary school counterparts when compared to non-ELLs. The researchers reminded that while some ELL students improve their English language proficiency and academic performance and thus exit LEP status, many students never exit LEP status and are considered long-term English learners.

Advances have been made in the study of ELLs' academic achievement in which more rigorous research designs were seen, yet there are still many shortcomings. Researchers were hampered by sample size limitations. Conflicting results were seen in part by use of test data from different measures, inconsistent application of LEP and ELL terminology, and selection of background variables and analysis techniques. These studies do not fully address the complexity of ELL students, and so this research is an earnest step toward demonstrating a few aspects of that diversity.

CHAPTER III

DATA AND METHODOLOGY

This research used data from one suburban school district in North Carolina. Information collected from an existing testing database includes math and reading scale scores. The first year in U.S. schools and LEP code fields from the existing database were used to create the number of years attending U.S. schools and LEP status variables for the study. Native language and birth country information were obtained from a separate district database, in which data had been input from Application for Enrollment forms. The form is sent home during the first week of classes, completed by the parent, and then returned to the school. No student background information was missing in this sample.

Sample

The participant data used for this research were not randomly selected due to the cohort design. The data include four consecutive years, from grade 4 to grade 7, of math and reading test scores on a common scale. To create the cohort, students were selected from grade 7 in the secondary database using one of these identifiers which classified a student as a current or former participant in ESL instruction: current LEP and reading exempt, current LEP, exited LEP within two years, or exited LEP identification. *Current LEP and reading exempt* students receive ESL instruction but they are also exempt from taking the End-of-Grade reading test because these students are in their first year in U.S. schools and earned scores below Level 4 Expanding on the reading subtest of an English language proficiency screener test, W-APT. *Current LEP* students receive ESL instruction and are not exempt from the End-of-Grade reading test. Test scores from students who *exited from LEP identification within the past two years* are designated as such because their scores are still included in the LEP subgroup for accountability purposes. *Exited LEP* students no longer qualify for ESL services due to their scores on the ACCESS for ELLs, an English language proficiency assessment, and are not considered LEP students for accountability purposes.

Students remained in this cohort if they had four years of test scores from grade 4 to grade 7 in both math and reading, resulting in two years of elementary and two years of middle school test score data while in attendance in the same school district. The final cohort has 373 students: 181 females and 192 males.

Dependent Variables

The state's mathematics and reading tests are multiple-choice assessments aligned with objectives from the state’s standard course of study. The tests are administered to students at the end of each school year. Multiple forms of each test are used. The test forms are parallel in terms of content coverage, and the tests are statistically equivalent at the total test score level. Scores from the tests determine yearly student performance, and previous and current year scores are used to calculate growth. A student’s raw score is converted to a scale score. Scores are reported as achievement levels, percentiles, and scale scores. These tests are high stakes, curriculum-based assessments, with the intent of holding students, staff, and schools accountable for academic performance. Educators use them to make promotion, remediation, acceleration, and graduation decisions, as well as program placement and course selections.

Math and reading developmental scale scores for grades four through seven were used in this research. Data include four consecutive years of math and reading scale scores; the years will be referred to as Grade 4 (G4), Grade 5 (G5), Grade 6 (G6), and Grade 7 (G7). The descriptive statistics for the cohort's math and reading scores are presented in Tables 3.1 and 3.2:

Table 3.1. Math Scale Scores

Grade	Min	Max	Mean	Std. Dev.
4th	326	373	348.94	8.94
5th	331	379	354.84	8.92
6th	334	382	355.72	9.03
7th	339	383	360.90	8.54

n = 373

Table 3.2. Reading Scale Scores

Grade	Min	Max	Mean	Std. Dev.
4th	319	366	339.94	9.28
5th	325	371	345.81	8.48
6th	328	371	350.62	8.49
7th	333	376	354.94	8.66

n = 373

Only regular, or first administration, scale scores were used in the analyses; therefore, retest scores were not included. During regular administrations, testing procedures are consistently applied to all test takers. However, retest administrations involve smaller groups, and testing procedures may be more informal and relaxed. Also, during the second test taking round, student attitudes may be altered due to fatigue or frustration. Students also participate in rigorous remediation sessions between regular and retest administrations which can also affect retest scores.

Student Variables

While rich, student-level data were available for this study, there were still challenges. For example, Spanish was the predominant native language and some of the groups had small sample sizes.

Native Language refers to the student's native, or first, language. Native language information was collected from the Primary/Home Language Survey section of the Application for Enrollment forms, whereby the parents noted the first language or dialect their child learned to speak. The information is used to initially identify students for possible ESL program services. Other than English, the five most frequently spoken languages in the school district are Spanish, Vietnamese, Arabic, Jarai, and Hindi. In this research, 36 of the district's more than 118 different languages were represented. Spanish (60.6%) was the largest native language group, then Vietnamese (10.7%), Arabic (4.3%), and Lao (2.4%).

While the pure intent of this research was to demonstrate the heterogeneity of ELLs and not to consider students as same, it became apparent that the numerous native languages in this sample would not allow for rigorous analysis if evaluated on their individual merits. Therefore, a purposeful and relevant grouping framework was created. Using a language framework for organization maintained the integrity of established language relationships and, as a result, allow for some demonstration of the linguistic diversity of ELL students. The initial step was to arrange each student's first language according to language family classification (Lewis, 2009; Aristar, 1990). The outcome of this arrangement was nine language families (see Table 3.3).

The Indo-European family represents the languages of 263 students, or 70.5 percent of the cohort. Due to its large representation in the sample, the Indo-European family was separated into two groups. The 226 Spanish-speaking students became one group and were eponymously titled. The 37 speakers from the remaining 15 Indo-European languages were placed in a second group termed Indo-European, in keeping with its language family classification. The Austro-Asiatic group, also named for its language family, includes 50 students who speak Khmer, Koho, or Vietnamese. At this stage of the grouping process, it was deemed statistically necessary to strategically cluster the remaining 17 languages. As a result, the language families were grouped according to the country where the language is prevalently spoken (Aristar, 1990). Using the geographic location by continent, the final two groups were titled African and Asian accordingly.

In summary, there are five *Native Language* groups: Spanish speakers in one group and the remaining 35 languages in four additional groups as displayed in Table 3.3:

Table 3.3. Native Language Groups

Language Family	Native Language (*n*)	Group Name	*n*
Afro-Asiatic Nilo-Saharan Nigero-Kordofanian Pidgin	Arabic (16) Somali (2) Amharic (1) Zarma (1) Ewe (1) Liberian English (2) Pidgin (2)	African	25
Korean Sino-Tibetan/Sino-Thai Austronesian	Korean (6) Lao (9) Chinese (5) Thai (2) Hmong (1) Jarai (5) Rade (3) Tagalog (2) Ponapean (1) Samoan (1)	Asian	35
Austro-Asiatic	Vietnamese (40) Khmer (8) Koho (2)	Austro-Asiatic	50
Indo-European	Hindi (7) French (4) Urdu (5) Farsi (3) Gujarti (3) Portuguese (3) Serbian (3) Italian (2) Afrikaans (1) Albanian (1) German (1) Nepali (1) Pashtu (1) Russian (1) Swedish (1)	Indo-European	37
Indo-European	Spanish (226)	Spanish	226

Country of Birth indicates the country where the student was born; this information was parent-reported on the Application for Enrollment form. Students from 35 countries were included in this research. The majority of students (57.6%) were born in the United States, followed by Mexico (20.4%) and Vietnam (5.9%). In specific reference to Spanish speakers, 62.4 percent were born in the U.S., 33.6 percent in Mexico, and the remaining 4.0 percent were born in Central, Latin, or South America. For analysis purposes, the country of birth variable was determined by geographic location and features three groups. The first group (n = 216) includes students born in the

U.S. or Canada. Students born in Mexico, Central America, South America, or the Caribbean are in the second group (*n* = 89). The third group (*n* = 68) includes students born in Africa, Asia, or Europe.

The native language and country of birth variables are displayed together in Table 3.4 to observe potential correlation issues. Three of the five language groups are split between two of three country groups. The Africa/Asia/Europe country group is uniformly divided between four of the five language groups. Although the Mexico/Central America/South America/Caribbean country group is almost exclusively attributed to the largest language group, Spanish speakers are still split between U.S./Canada (62.4%) and Mexico/Central America/South America/Caribbean (37.6%) for country of birth group.

Table 3.4. Native Language Group by Country of Birth Group

		Country of Birth Group		
		Africa/Asia/ Europe	**Mexico/CA/SA /Caribbean**	**U.S./Canada**
Native Language Group	**African**	15	1	9
	Asian	18	0	17
	Austro-Asiatic	17	0	33
	Indo-European	18	3	16
	Spanish	0	85	141
	Total	68	89	216

The *Number of Years that the student had attended school in the United States* at the end of grade 4 is included as a continuous variable of interest. The first year in U.S. schools was entered by parents on the Primary/Home Language Survey section of the Application for Enrollment form and was used to create this variable. The majority of students started their U.S. schooling in Kindergarten; however, 14 percent entered the system at 2nd grade or later. The number of years at Time Point 1, or at initial status, that students attended U.S. schools is shown in Table 3.5. One year means that the student's 4th grade year was the first year attending a U.S. school. Six years indicates that the student began schooling in the U.S. in Pre-Kindergarten or had been retained one grade prior to 4th grade. The average number of years these students had attended U.S. schools is 4.47.

Table 3.5. Number of Years Attending U.S. Schools (Time Point 1)

Years	Number	Percent
1	8	2.1
2	14	3.8
3	30	8.0
4	73	19.6
5	239	64.1
6	9	2.4

The average number of years attending U.S. schools is presented in Table 3.6 for the five native language groups. The Asian native language students possessed the lowest average (3.91 years). The Austro-Asiatic group had attended U.S. schools for the longest period of time (4.80 years), which is nearly one year more than students in the Asian native language group.

Table 3.6. Native Language Group by Average Number of Years in U.S. Schools

		Average Number of Years in U.S. Schools
Native Language Group	**African**	4.04
	Asian	3.91
	Austro-Asiatic	4.80
	Indo-European	4.46
	Spanish	4.53
	Total	4.47

LEP Status refers to whether the student still qualified for ESL services in 4th grade, as determined by scores on the W-APT, and was still identified as LEP. In this school district's ESL program model, English is the language of instruction while the student's native language is used for clarification purposes only. In the elementary grades, ELLs leave regular classroom instruction during the language arts block, for example, and attend ESL class. In middle school, students attend ESL instruction as a regular class period. When an ELL student no longer qualifies for services due to scores on the ACCESS for ELLs, the student's status is stated as exited from LEP identification. For this research, the student's LEP status at grade 4 was considered. Existing information from LEP fields in the database was used to create this variable. The data are coded as follows:

0 = *Exited*
The student scored Level 5.0 Bridging in all four subtests of the WIDA-ACCESS for ELLs and exited LEP identification. This student does not participate in the ESL program because he is no longer identified as LEP (n = 99).

1 = *Current*
The student is classified as an LEP student based on his performance on the WIDA-ACCESS Placement Test (W-APT). This student participates in the ESL program (n = 274).

Hierarchical Linear Modeling Approach

The analysis of longitudinal data can focus on changes in mean responses over time and on the relationship of those changes to predictor variables, so hierarchical linear modeling is a useful choice to address the research questions posed here. Because this study featured a multiple-time-point design with small group sample sizes and test scores nested within individual students, using a hierarchical linear model offered statistical precision and an integrated approach to examine predictors of individual growth (Raudenbush & Bryk, 2002). A linear growth model was specified where initial status and individual growth rates were estimated. In that the math and reading test data are measured on a common metric, potential change across the four time periods will reflect growth and not a change in the measurement scale (Bryk & Raudenbush, 1987). The intercept and slope for each student was estimated at level 1. Then, at level 2, the intercept and slope was modeled as a function of multiple student background variables. The intent was for this hierarchical linear modeling approach to highlight the variability that exists between ELL students which, in turn, would help extricate information about individual predictors that affect academic achievement.

A model-building approach was used whereby level-2 predictors were added to the level-1 model in a stepwise process based on empirical and theoretical considerations. Model estimates were computed using HLM 7 software (Raudenbush et al., 2011).

The model parameters were estimated using the full maximum likelihood (FML) method. FML was selected over restricted maximum likelihood (REML), the HLM7 default, because of the deviance values that the FML method provides. The values from the FML method were used in deviance tests to help determine if an alternative model was a better fit than a previous model. The REML method only compares random effect changes which would not have provided sufficient information for use in model fit evaluation (Snijders & Bosker, 1999).

The dependent and student background variables were not centered in the model specification process. The variables were left uncentered because the data feature equal time points across students and the intercept should be interpreted from initial status. In addition, dummy-coded variables were used to determine the difference between the respective groups on the outcome, and to do so, the variables should be left uncentered.

Spanish language is the reference group for the Native Language group variable. The four remaining language groups were compared to Spanish because Spanish is often used as the sample ELL or native language group and, as a result, is heavily researched. For the Country of Birth variable, the United States and Canada group is the reference group because it has the greatest number of students in the sample (57.6%). Student-level variables are summarized in Table 3.7 with identifying name abbreviations and descriptive statistics:

Table 3.7. Level-2 Variables Key

Variable	Name	Values	P/M
Language group African	AFR	0 = non-African language group 1 = African language group	0.07
Language group Asian	ASN	0 = non-Asian language group 1 = Asian language group	0.09
Language group Austro-Asiatic	AUAS	0 = non-Austro-Asiatic language group 1 = Austro-Asiatic language group	0.13
Language group Indo-European	INDO	0 = non-Indo-European language group 1 = Indo-European language group	0.10
Birth country in Mexico, Central America, South America, Caribbean	MCSC	0 = not Mexico, C. America, S. America, Carib 1 = Mexico, C. America, S. America, Carib	0.24
Birth country in Africa, Asia, Europe	AAE	0 = not Africa, Asia, Europe 1 = African, Asia, Europe	0.18
LEP status	LEP	0 = exited LEP, 1 = current LEP	0.74
Number of years attending U.S. schools	YRS	1 to 6 years	4.5

At level 1, this linear individual growth model designates academic achievement at time *t* of student *i*, indicating time is nested within students with the assumption that the growth parameters vary across individuals.

$$Y_{ti} = \pi_{0i} + \pi_{1i}a_{ti} + e_{ti} \tag{1}$$

where

Y_{ti} represents the domain score at time *t* for each student *i*

π_{0i} represents the initial status of student *i*, representing initial domain outcome for student *i* in grade 4

π_{1i} represents the growth rate and expected change for student *i* from grade 4 to grade 7

a_{ti} represents the test occasions at grades 4, 5, 6, and 7 (coded 0, 1, 2, 3) at time *t* for student *i*

e_{ti} represents the random error/residual variance at time *t* for student *i*

Equation 2 denotes level 2 as an unconditional model where the level-2 predictors had yet to be introduced. This unconditional model provided baseline estimates for evaluating the prospective level-2 models.

$$\pi_{0i} = \beta_{00} + r_{0i},$$
$$\pi_{1i} = \beta_{10} + r_{1i} \qquad (2)$$

where

β_{00} represents the intercept that indicates initial domain score

β_{10} represents the linear growth rate

r_{0i} represents the initial status random error at time *t* for student *i*

r_{1i} represents the growth rates random error at time *t* for student *i*

The models were intended to provide information as to whether initial status and growth rates were functions of specific student characteristics. Both the intercept and slope were allowed to vary (Bryk & Raudenbush, 1987). The level-1 model remained unchanged, while the following predictors were entered one by one to sequentially build up the level-2 model: LEP Status, Native Language groups, Number of Years Attending U.S. Schools, and Country of Birth groups. These level-2 models examined the extent that the background variables were predictive of academic achievement. Predictors remained in the model if found to make a significant contribution after their addition.

The LEP Status variable was the first predictor entered into the level-2 model, as specified by Equation 3. LEP status has been a widely-used variable of interest by researchers, yet it has been studied more in terms of English Language Learners' academic performance when compared to native English speakers (Roberts et al., 2010; Kieffer, 2008; Sung & Chang, 2008; Lesaux et al., 2007; D'Angiulli et al., 2004; Hofstetter, 2003; Abedi & Lord, 2001). LEP status has also been occasionally examined as differences between current and exited LEP students on a longitudinal basis (Chen, 2010; deJong, 2004; Wang & Goldschmidt, 1999). This model determined the extent to which initial LEP status, whether no longer receiving or currently receiving ESL instruction at grade 4, was related to achievement over the four year period.

$$\pi_{0i} = \beta_{00} + \beta_{01} (LEP)_i + r_{0i} ,$$
$$\pi_{1i} = \beta_{10} + \beta_{11} (LEP)_i + r_{1i}. \quad (3)$$

where

β_{01} represents the extent to which initial status of LEP students is different from exited LEP students

β_{11} represents the extent to which growth rates of LEP students are different from exited LEP students

Four native language groups were added next as shown in Equation 4, while controlling for LEP Status. The native language variable has been included in a modest number of studies, but the focus has generally been on Spanish speakers. These predictors helped determine how speaking a language different from the majority of ELL students affected academic achievement. To that end, the Spanish language group served as the reference group.

$$\pi_{0i} = \beta_{00} + \beta_{01}\,(LEP)_i + \beta_{02}\,(AFR)_i + \beta_{03}(ASN)_i + \beta_{04}(AUAS)_i + \beta_{05}(INDO)_i + r_{0i}\,,$$

$$\pi_{1i} = \beta_{10} + \beta_{11}\,(LEP)_i + \beta_{12}\,(AFR)_i + \beta_{13}(ASN)_i + \beta_{14}(AUAS)_i + \beta_{15}(INDO)_i + r_{1i} \qquad (4)$$

where

β_{02} represents the extent to which initial status of African language group students is different from Spanish language group students, while controlling for LEP status

β_{03} represents the extent to which initial status of Asian language group students is different from Spanish language group, while controlling for LEP status

β_{04} represents the extent to which initial status of Austro-Asiatic language group students is different from Spanish language group students, while controlling for LEP status

β_{05} represents the extent to which initial status of Indo-European language group students is different from Spanish language group students, while controlling for LEP status

β_{12} represents the extent to which growth rates of African language group students are different from Spanish language group students, while controlling for LEP status

β_{13} represents the extent to which growth rates of Asian language group students are different from Spanish language group students, while controlling for LEP status

β_{14} represents the extent to which growth rates of Austro-Asiatic language group students are different from Spanish language group students, while controlling for LEP status

β_{15} represents the extent to which growth rates of Indo-European language group students are different from Spanish language group students, while controlling for LEP status

The Number of Years Attending U.S. Schools, a continuous variable, was considered next (Equation 5). This background variable has often been framed as age at immigration in the few research studies it has been featured. This aspect of an ELL student's life is under-researched so new findings could provide additional information about these students' educational journeys.

$$\pi_{0i} = \beta_{00} + \beta_{01}(LEP)_i + \beta_{02}(AFR)_i + \beta_{03}(ASN)_i + \beta_{04}(AUAS)_i + \beta_{05}(INDO)_i + \beta_{06}(YRS)_i + r_{0i},$$

$$\pi_{1i} = \beta_{10} + \beta_{11}(LEP)_i + \beta_{12}(AFR)_i + \beta_{13}(ASN)_i + \beta_{14}(AUAS)_i + \beta_{15}(INDO)_i + \beta_{16}(YRS)_i + r_{1i} \quad (5)$$

where

β_{06} represents the extent to which the number of years students attended U.S. schools affects initial status, while controlling for LEP status and native language

β_{16} represents the extent to which the number of years students attended U.S. schools affects growth rate, while controlling for LEP status and native language

The Country of Birth group variables were the final predictors introduced (Equation 6), which included two country of birth groups that were compared to the predominant United States/Canada group.

$$\pi_{0i} = \beta_{00} + \beta_{01}(LEP)_i + \beta_{02}(AFR)_i + \beta_{03}(ASN)_i + \beta_{04}(AUAS)_i + \beta_{05}(INDO)_i$$
$$+ \beta_{06}(YRS)_i + \beta_{07}(MCSC)_i + \beta_{08}(AAE)_i + r_{0i},$$
$$\pi_{1i} = \beta_{10} + \beta_{11}(LEP)_i + \beta_{12}(AFR)_i + \beta_{13}(ASN)_i + \beta_{14}(AUAS)_i + \beta_{15}(INDO)_i$$
$$+ \beta_{16}(YRS)_i + \beta_{17}(MCSC)_i + \beta_{18}(AAE)_i + r_{1i} \quad (6)$$

where

β_{07} represents the extent to which initial status of students born in Mexico, Central America, South America, or the Caribbean is different from students born in the U.S. or Canada, while controlling for LEP status, native language, and years in U.S. schools

β_{08} represents the extent to which initial status of students born on the continents of Africa, Asia, or Europe is different from students born in the U.S. or Canada, while controlling for LEP status, native language, and years in U.S. schools

β_{17} represents the extent to which growth rates of students born in Mexico, Central America, South America, or the Caribbean are different from students born in the U.S. or Canada, while controlling for LEP status, native language, and years in U.S. schools

β_{18} represents the extent to which growth rates of students born on the continents of Africa, Asia, or Europe are different from students born in the U.S. or Canada, while controlling for LEP status, native language, and years in U.S. schools

Adequacy of the model specification was determined by assessing multilevel model assumptions. The following two assumptions were checked: (1) predictors were independent of errors between students; and (2) errors were assumed to be normally distributed with mean 0 and variance σ^2. The Level-1 residuals were used in the diagnostic techniques to assess these assumptions. To evaluate normality and dependency, residuals were checked via Q-Q plots and histograms. To check for heterogeneity, the homogeneity of Level-1 variances was tested using a χ^2 statistic. These results helped determine if scores vary across students.

The intent of this research was to emphasize the diversity of ELL students. The first research purpose was to examine how initial LEP status impacted academic achievement of these same students. The second purpose was to explore how level-2 predictors, including native language, country of birth, and number of years attending U.S. schools, impacted reading and math. Model comparisons were made based on deviance test results and changes in coefficient significance. The deviance statistic was computed for each model in which the deviance can be viewed as a measure of model fit (Raudenbush & Bryk, 2002). The deviance difference between the two models is distributed as a chi-square statistic, and helped determine if one model was a significantly better fit to the data than the subsequent model. To begin, the first alternative model was compared to the unconditional model to determine if it was an improvement. Accordingly, each subsequent model was compared with the previous model. If a level-2 predictor was not an improvement, then it was considered for removal from the remaining models. While deviance statistics were used to help determine model fit, the coefficients were also observed for significance to help determine if the addition of variables was more informative.

Using hypothesis tests for fixed effects, the ratio of the estimated effects to their standard error for mean initial status (β_{00}) and mean growth rate (β_{10}) was examined. Learning which student background variables had an effect on academic achievement was determined by the significance of the coefficient at level 2. For example, this test reflected if mean scores from the Asian language group differed from the Spanish reference group at initial status. Additionally, hypothesis tests signified if being born outside the U.S. or Canada, the number of years attending U.S. schools, or being classified as an LEP student in fourth grade had an effect on math and reading performance.

CHAPTER IV

RESULTS

The purpose of this research was to emphasize the diversity of ELL students by examining how initial LEP status impacts academic achievement and by exploring the relationship of student variables to academic performance in reading and math through the use of hierarchical linear modeling. The research questions were:

1. How does 4^{th} grade LEP Status relate to academic performance in math and reading at initial status and over the following four years?
2. In what ways do native language, country of birth, and number of years attending U.S. schools predict math and reading achievement of elementary and middle school ELL students?

Based on the availability of data, this research featured multiple variables including math and reading scores (four time points), LEP Status (current or exited), Native Language groups (African, Asian, Austro-Asiatic, Indo-European, and Spanish), Number of Years Attending U.S. Schools, and Country of Birth groups (U.S./Canada, Mexico/Central America/South America/Caribbean, and Africa/Asia/Europe).

Process for Model Evaluation

The multilevel model assumptions were assessed to consider adequacy of the model specifications. The residual file, which holds the difference between the fitted and observed values for each level-1 unit, was used to examine the assumption of normal distribution of level-1 errors.

The final hierarchical linear model was determined by using a model building process, whereby student background variables were added one by one to the unconditional models for math and reading. This sequential approach is reflected in both the analyses and the research questions. The unique contribution of each predictor, independent of other predictors in the model, was reviewed. Model comparisons were made using the deviance statistic computed for the model runs and by observing the significance levels of the coefficients. The difference between the deviance statistics of the current and each alternative model was considered in terms of model fit. If the comparison test result and the new predictor's coefficient were not significant at $p < .05$, then the alternative model might not be a better fit. The alternative model was also evaluated by observing the coefficients for significance, as well as assessing previous analyses and theoretical considerations.

Math – Model Evaluation

The Q-Q plot (Appendix B) shows the level-1 math residuals based on the unconditional model. The plot formed a straight diagonal line indicating the errors conformed closely to normality. The distribution (Appendix B) was reasonably smooth and symmetrical. This visual inspection supported that the normality assumption held for the math score data, and the error terms were normally distributed.

A test of homogeneity of level-1 variance was run to determine if the math scores varied across students. The result, χ^2 = 417.57 with df = 370 and p = 0.04, suggested that the residual variances did differ significantly, which refuted the assumption of homogeneity of the level-1 variances. This result indicated there was variability among the level-2 units in terms of the residual within-student variance (Raudenbush et al., 2011); therefore, adding level-2 predictors to examine the variability would be appropriate.

For the unconditional model, math achievement was modeled without predictors and was allowed to vary across students. Results from the analysis of the unconditional model are displayed in Table 4.1. The Level-1 variance (σ^2) was 15.17 and the Level-2 variance (r_{0i}) was 67.55, which produced an intraclass correlation coefficient (ICC) of 0.82. The magnitude of this ICC suggested that 82 percent of the variance in math achievement was due to within-student differences, and that 18 percent of the variance can be attributed to between-student differences, further confirming that hierarchical linear modeling was appropriate for telling a more informative story about the math data.

The fixed effects with robust standard errors were reviewed. Using robust standard errors is appropriate in these analyses because the sample featured a moderate to large number of level-2 units. Even if the HLM assumptions were incorrect, the standard errors would be consistent (Raudenbush & Bryk, 2002). The estimated mean intercept for math achievement was significant (β_{00} = 349.58, SE = 0.46, p < 0.01) and represented the mean math initial status (see Table 4.1). The mean growth rate, β_{10}, was 3.68 (SE = 0.10, $p < 0.01$). The significance of the slope indicated that students demonstrated, on average, 3.68 scale score points of growth in math each year. Based on the slope's estimated standard deviation of 0.85 (see Table 4.1) and the mean growth rate, it would be expected that a student whose growth was one standard deviation above average would experience growth at a rate of 4.53 points (0.85 + 3.68) each year.

The variance of the intercept (r_{0i}) was statistically significant ($\chi^2 = 2745.11$, $df = 372$, $p < 0.01$). This result signified there was variability in math achievement at initial status. The variance of the slope (r_{1i}), was also significant ($\chi^2 = 460.94$, $df = 372$, $p = 0.00$) which implied there were individual differences among growth rates. The intercept was quite reliable ($\pi_{0i} = 0.86$). Such strong reliability provided more evidence toward examining true individual differences in initial status for math. In contrast, the slope had a much lower reliability estimate ($\pi_{1i} = 0.19$) which suggested that the math growth rates varied only slightly within students.

Table 4.1. Unconditional Model – Math

Fixed Effect	Coefficient	SE	t Ratio	df	*p*-value
Mean initial status, β_{00}	349.58	0.46	763.65	372	<0.01
Mean growth rate, β_{10}	3.68	0.10	36.68	372	<0.01

Random Effect	SD	Variance Component	df	χ^2	*p*-value
Initial status, r_{0i}	8.22	67.55	372	2745.11	<0.01
Growth rate, r_{1i}	0.85	0.72	372	460.94	<0.01
Level-1 error, e_{ti}	3.90	15.17			

Reliability of OLS Regression Coefficient Estimate	
Initial status, π_{0i}	0.86
Growth rate, π_{1i}	0.19

In the next step of the model building process, LEP status was introduced as a predictor at Level-2 to determine differences in math for students who were or were not classified as Limited English Proficient in 4th grade. The deviance values of this model (D = 9394.72) and the unconditional model (D = 9440.97) were used in the model comparison test, and produced a χ^2 statistic of 46.25 (df = 2, p < 0.01). This result indicated that Model 1 was significantly different from the unconditional model and the addition of LEP Status may be an improvement. Accordingly, the intercept coefficient for LEP Status (see Table 4.2) was found significant (β_{01} = -6.53, SE = 1.05, p < 0.01). The slope, however, was not significant (β_{11} = 0.14, SE = 0.23, p = 0.55). Based on the deviance statistic comparisons and the significance at initial status, LEP Status remained in the model. In this first model, LEP Status predicted lower initial math scores but was not associated with different rates of growth.

Table 4.2. Model 1 - Math (Addition of LEP Status)

Fixed Effect	Coefficient	SE	t Ratio	df	*p*-value
Model for initial status, π_{0i}					
Initial math score, β_{00}	354.38	0.93	382.33	371	<0.01
LEP status, β_{01}	-6.53	1.05	-6.25	371	<0.01
Model for growth rate, π_{1i}					
Initial growth rate, β_{10}	3.58	0.19	18.44	371	<0.01
LEP status, β_{11}	0.14	0.23	0.60	371	0.55

Random Effect	SD	Variance Component	df	χ^2	*p*-value
Initial status, r_{0i}	7.70	59.22	371	2452.88	<0.01
Growth rate, r_{1i}	0.84	0.71	371	460.50	<0.01
Level-1 error, e_{ti}	3.90	15.17			

In Model 2, four Native Language group variables were added to learn if a student's first language affected academic achievement in math. The deviance values of Model 2 (D = 9376.19) and Model 1 (D = 9394.72) were used to compare model fit, which produced a χ^2 statistic of 18.53 (df = 8, p = 0.02). This outcome suggested that Model 2 was significantly different from Model 1. As a result, the inclusion of the Native Language groups as predictors for academic achievement in math may result in a more informative model.

Two of the four native language group intercept coefficients were significant (see Table 4.3). The Asian group (β_{03} = 3.99, SE = 1.51, p = 0.01), of which 43 percent spoke Korean or Lao, and the Austro-Asiatic group (β_{04} = 2.50, SE = 1.15, p = 0.03), with 80 percent speaking Vietnamese, were both significantly different from the Spanish language group at initial status. The coefficients for the African and Indo-European native language groups were not significant at initial status. LEP Status remained significant (β_{01} = -6.57, SE = 1.02, $p < 0.01$) in Model 2. None of the four language groups demonstrated a growth rate significantly different from the Spanish group (all $p >$ 0.05). Based on the deviance statistic comparisons and the significance at initial status for LEP Status and two of the four language groups, the Native Language variables were retained in the model. Although not associated with different rates of growth, Native Language group did predict higher initial math scores.

Table 4.3. Model 2 - Math (Addition of Native Language Groups)

Fixed Effect	Coefficient	SE	t Ratio	df	*p*-value
Model for initial status, π_{0i}					
Initial math score, β_{00}	353.65	1.00	353.41	367	<0.01
LEP status, β_{01}	-6.57	1.02	-6.41	367	<0.01
African language group, β_{02}	-0.30	1.93	-0.16	367	0.88
Asian language group, β_{03}	3.99	1.61	2.48	367	0.01
Austro-Asiatic language group, β_{04}	2.50	1.15	2.17	367	0.03
Indo-European language group, β_{05}	0.71	1.54	0.46	367	0.65
Model for growth rate, π_{1i}					
Initial growth rate, β_{10}	3.47	0.23	15.35	367	<0.01
LEP status, β_{11}	0.14	0.23	0.61	367	0.54
African language group, β_{12}	-0.08	0.49	-0.16	367	0.87
Asian language group, β_{13}	0.35	0.36	0.97	367	0.33
Austro-Asiatic language group, β_{14}	0.26	0.25	1.04	367	0.30
Indo-European language group, β_{15}	0.39	0.35	1.10	367	0.27

Random Effect	SD	Variance Component	df	χ^2	*p*-value
Initial status, r_{0i}	7.58	57.41	367	2389.30	<0.01
Growth rate, r_{1i}	0.83	0.69	367	457.31	<0.01
Level-1 error, e_{ti}	3.90	15.17			

Next, the Number of Years Attending U.S. Schools variable was included in Model 3. The deviance values of Model 3 (D = 9362.09) and Model 2 (D = 9376.19) were used in the model comparison test. The result produced a χ^2 statistic of 14.10 (df = 2, $p < 0.01$), which indicated that Model 3 may be an improvement over Model 2. As shown in Table 4.4, the coefficient for the Number of Years Attending U.S. Schools variable was significant at initial status (β_{06} = -1.49, SE = 0.49, $p < 0.01$), but not significant for growth (β_{16} = -0.07, SE = 0.12, p = 0.59). The Asian (β_{03} = 3.09, SE = 1.55, $p < 0.05$) and Austro-Asiatic (β_{04} = 2.87, SE = 1.16, p = 0.01) native language groups were still significant at initial status, as was LEP Status (β_{01} = -7.55, SE = 1.05, $p < 0.01$). The Number of Years Attending U.S. Schools variable was determined to be an informative improvement in the model and, thus, it was kept, together with LEP Status and the four Native Language groups, in the subsequent model.

Table 4.4. Model 3 - Math (Addition of Years in U.S. Schools)

Fixed Effect	Coefficient	SE	t Ratio	df	*p*-value
Model for initial status, π_{0i}					
Initial math score, β_{00}	359.65	2.09	172.42	366	<0.01
LEP status, β_{01}	-7.55	1.05	-7.23	366	<0.01
African language group, β_{02}	-1.14	2.01	-0.57	366	0.57
Asian language group, β_{03}	3.09	1.55	2.00	366	<0.05
Austro-Asiatic language group, β_{04}	2.87	1.16	2.47	366	0.01
Indo-European language group, β_{05}	0.53	1.52	0.35	366	0.73
Years in U.S. Schools, β_{06}	-1.49	0.47	-3.18	366	<0.01
Model for growth rate, π_{1i}					
Initial growth rate, β_{10}	3.74	0.55	6.75	366	<0.01
LEP status, β_{11}	0.10	0.24	0.40	366	0.69
African language group, β_{12}	-0.12	0.50	-0.23	366	0.82
Asian language group, β_{13}	0.31	0.37	0.83	366	0.41
Austro-Asiatic language group, β_{14}	0.28	0.25	1.11	366	0.27
Indo-European language group, β_{15}	0.38	0.36	1.07	366	0.29
Years in U.S. Schools, β_{16}	-0.07	0.12	-0.54	366	0.59

Random Effect	SD	Variance Component	df	χ^2	*p*-value
Initial status, r_{0i}	7.46	55.65	366	2327.23	<0.01
Growth rate, r_{1i}	0.83	0.68	366	456.89	<0.01
Level-1 error, e_{ti}	3.90	15.17			

Lastly, the Country of Birth group variables were considered in Model 4. The deviance values for Model 4 (D = 9357.58) and Model 3 (D = 9362.09) were used to compare model fit, which produced a χ^2 statistic of 4.51 (df = 4, p = 0.34). This result suggested that Model 4 may not be an improvement in fit compared to Model 3. The coefficients were then observed for significant values to help determine the relevance of including the Country of Birth groups in the final model.

The Mexico/Central America/South America/Caribbean country of birth group was significantly different at initial status (β_{08} = -2.18, SE = 1.06, p = 0.04) from the

U.S./Canada reference group; but, the Africa/Asia/Europe group coefficient was not significant ((β_{07} = -1.57, SE = 1.42, p = 0.27), as shown in Table 4.5. And, like the previously-entered variables, the two Country of Birth groups did not show a growth rate significantly different from their reference group (all p > .05). The LEP Status (β_{01} = -7.49, SE = 1.05, p < 0.01), Austro-Asiatic native language group (β_{04} = 2.68, SE = 1.29, p = 0.04) and Years Attending U.S. Schools (β_{06} = -1.85, SE = 0.48, p < 0.01) predictors were still significant at initial status. However, in this model, the Asian native language group (β_{03} = 2.86, SE = 1.76, p = 0.11) dropped below the significance level of p < 0.05. It is possible this change in significance could be an interaction between native language and country of birth. In this cohort, no Asian native language groups students were represented in the Mexico/Central America/South America/Caribbean country of birth group (see Table 3.4), so the lack of significance could be a product of the sample itself. Based on the significance at initial status for one of the two Country of Birth predictors, Country of Birth was retained in the final math model in addition to all previously-entered variables (see Table 4.5).

Table 4.5. Final Model / Model 4 - Math (Addition of Country of Birth Groups)

Fixed Effect	Coefficient	SE	t Ratio	df	*p* -value
Model for initial status, π_{0i}					
Initial math score, β_{00}	361.68	2.25	160.98	364	<0.01
LEP status, β_{01}	-7.49	1.05	-7.13	364	<0.01
African language group, β_{02}	-1.10	2.04	-0.54	364	0.59
Asian language group, β_{03}	2.86	1.76	1.62	364	0.11
Austro-Asiatic language group, β_{04}	2.68	1.29	2.08	364	0.04
Indo-European language group, β_{05}	0.63	1.76	0.36	364	0.72
Years in U.S. Schools, β_{06}	-1.85	0.48	-3.87	364	<0.01
Africa/Asia/Europe group, β_{07}	-1.57	1.42	-1.10	364	0.27
Mexico/CA/SA/Carib group, β_{08}	-2.18	1.06	-2.06	364	0.04
Model for growth rate, π_{1i}					
Initial growth rate, β_{10}	3.53	0.58	6.11	364	<0.01
LEP status, β_{11}	0.09	0.24	0.37	364	0.71
African language group, β_{12}	-0.12	0.49	-0.24	364	0.81
Asian language group, β_{13}	0.37	0.41	0.82	364	0.41
Austro-Asiatic language group, β_{14}	0.30	0.30	1.00	364	0.32
Indo-European language group, β_{15}	0.38	0.37	1.00	364	0.32
Years in U.S. Schools, β_{16}	-0.03	0.13	-0.23	364	0.82
Africa/Asia/Europe group, β_{17}	0.15	0.32	0.49	364	0.63
Mexico/CA/SA/Carib group, β_{18}	0.22	0.27	0.82	364	0.41

Random Effect	SD	Variance Component	df	χ^2	*p* -value
Initial status, r_{0i}	7.41	54.85	364	2299.28	<0.01
Growth rate, r_{1i}	0.82	0.67	364	455.88	<0.01
Level-1 error, e_{ti}	3.90	15.17			

Math - Summary

The estimated fixed effects of the final math model are presented in Table 4.5. Eight predictors were included in the final model: initial LEP Status, four Native Language groups, Number of Years in U.S. Schools, and two Country of Birth groups. Results of the final model showed that LEP Status was significantly related to math scores at initial status (β_{01} = -7.49, SE = 1.05, p < 0.01). The scores for current LEP students were 7.49 points lower than those of exited LEP students at initial status (4th grade). The difference was close to one standard deviation for this math assessment (SD = 8.94), producing a fairly large effect of .84. The coefficient for the slope was not significant, which indicated that LEP Status did not have an impact on rates of growth in math.

Two of the four native language groups were significant predictors of math scores at initial status. The coefficient for the Austro-Asiatic language group (β_{04} = 2.68, SE = 1.29, *p* = 0.04) was significant. This finding indicated that students in the Austro-Asiatic native language group scored an average of 2.68 points higher than the Spanish language group students, the reference group, in 4th grade. The Asian language group coefficient was deemed significant in Model 2 (β_{03} = 3.99, SE = 1.51, *p* = 0.01) and Model 3 (β_{03} = 3.09, SE = 1.55, $p < 0.05$), but dropped below significance in Model 4, the final model. Again, the coefficients for the slope were not significant, which showed that native language had no effect on math growth rates when compared with the Spanish language group.

The number of years attending U.S. schools was a significant predictor at initial status in math (β_{06} = -1.49, SE = 0.49, $p < 0.01$). This result indicated that, for each additional year a student had attended U.S. schools, the student earned 1.49 scale score points fewer in math in 4th grade. It seems counterintuitive that the longer ELL students were in school that their academic performance declined; however, this finding, too, could be a product of the sample. In the reference group, data were sparse in that only two percent of students had been in U.S. schools for one year at initial status (see Table

3.5), compared to 64 percent who had attended U.S. schools for five years. The final model showed that the Mexico/Central America/South America/Caribbean country of birth group had significantly lower initial math scores (β_{08} = -2.18, SE = 1.06, p = 0.04) than the U.S./Canada reference group. Therefore, students in the Mexico/Central America/South America/Caribbean country of birth group earned an average of 2.18 scale score points lower in 4th grade than students in the U.S./Canada group. The Country of Birth variable was not significantly related to math growth rates.

Reading – Model Evaluation

The Q-Q plot (Appendix C) shows the level-1 reading residuals based on the unconditional model. Although the tails were slightly curved, the plot was fairly linear which suggested the errors were normally distributed. The dots at the left and right of the graph veered away from the line, indicating there was some divergence from normality in the tails of the distribution. The histogram (Appendix C) of the level-1 reading residuals displayed a relatively normal distribution with some extreme values at the center. These data were accurate and represented the students' test score achievement, so any deviations were considered small enough to allow interpretation of results.

A test of homogeneity of level-1 variance was run to determine if the reading scores varied across students. The result, χ^2 = 385.97, with 370 df, and p = 0.27, suggested that the residual variances did not differ significantly, which confirmed the assumption of homogeneity of the level-1 variances. This result indicated that there was no variability among the level-2 units in terms of the residual within-student variance (Raudenbush et al., 2011).

For the unconditional model, reading achievement was modeled without predictors and was allowed to vary across students. Results from the analysis of the unconditional model are displayed in Table 4.6. The Level-1 variance (σ^2) was 12.24 and the Level-2 variance (r_{0i}) was 69.63, which produced an intraclass correlation coefficient (ICC) of 0.85. The magnitude of this ICC suggested that 85 percent of the variance in reading achievement was due to within-student differences, and that 15 percent of the variance can be attributed to between-student differences, which offered further confirmation that hierarchical linear modeling was appropriate for the reading data.

The coefficients for both the intercept and slope were significant. The estimated mean intercept, β_{00}, for reading achievement was 340.35 (SE = 0.46, $p < 0.01$), and represented the average reading score at initial status. The mean growth rate coefficient, β_{10}, was 4.98 (SE = 0.10, $p < 0.01$). This result signified that, on average, students demonstrated 4.98 scale score points of growth in reading each year. Based on the estimated slope's standard deviation 1.18 (see Table 4.6) and the mean growth rate (4.98), it would be expected that a student whose growth was one standard deviation above average would experience growth at a rate of 6.16 points each year.

The variance of the intercept (r_{0i}) was statistically significant (χ^2 = 3404.72, df = 372, $p < 0.01$). This result implied variability in reading achievement. As for the slope (r_{1i}), the variance was also significant (χ^2 = 586.71, df = 372, $p < 0.01$); this significance suggested there were individual differences among reading growth rates. The estimated reliability for initial status in reading was 0.89. This high reliability estimate offered strong support toward true individual differences in reading. The slope had a lower reliability estimate (π_{1i} = 0.36); however, which indicated the reading growth rates might be limited in variability within students.

Table 4.6. Unconditional Model - Reading

Fixed Effect	Coefficient	SE	t Ratio	df	*p*-value
Mean initial status, β_{00}	340.35	0.46	743.32	372	<0.01
Mean growth rate, β_{10}	4.98	0.10	49.02	372	<0.01

Random Effect	SD	Variance Component	df	χ^2	*p*-value
Initial status, r_{0i}	8.34	69.63	372	3404.72	<0.01
Growth rate, r_{1i}	1.18	1.40	372	586.71	<0.01
Level-1 error, e_{ti}	3.50	12.24			

Reliability of OLS Regression Coefficient Estimate	
Initial status, π_{0i}	0.89
Growth rate, π_{1i}	0.36

Initial LEP status was introduced in the first alternative model as a predictor at Level-2 to determine differences in students who were or were not classified as Limited English Proficient in 4th grade. The deviance values of Model 1 (*D* = 9210.79) and the unconditional model (*D* = 9278.46) were used in the model comparison test, and produced a χ^2 statistic of 67.67 (df = 2, p < 0.01). This result indicated that Model 1 was significantly different from the unconditional model and that the addition of the LEP Status predictor might be an improvement. The intercept coefficient for LEP Status (see Table 4.7) was significant (β_{01} = -7.86, SE = 0.91, p < 0.01); however, the slope's intercept was not (β_{11} = 0.28, SE = 0.22, *p* = 0.21). Based on the deviance test result and the significance at initial status, LEP Status was retained in the model. In Model 1, LEP Status predicted lower initial scores but was not associated with different rates of growth.

Table 4.7. Model 1 - Reading (Addition of LEP Status)

Fixed Effect	Coefficient	SE	t Ratio	df	*p* -value
Model for initial status, π_{0i}					
Initial reading score, β_{00}	346.13	0.75	459.63	371	<0.01
LEP status, β_{01}	-7.86	0.91	-8.67	371	<0.01
Model for growth rate, π_{1i}					
Initial growth rate, β_{10}	4.77	0.19	25.66	371	<0.01
LEP status, β_{11}	0.28	0.22	1.27	371	0.21

Random Effect	SD	Variance Component	df	χ^2	*p* -value
Initial status, r_{0i}	7.59	57.59	371	2880.48	<0.01
Growth rate, r_{1i}	1.18	1.39	371	584.36	<0.01
Level-1 error, e_{ti}	3.50	12.24			

In Model 2, the Native Language group variables were added to learn if a students' first language affected academic achievement in reading. The deviance values of Model 2 (*D* = 9196.67) and Model 1 (*D* = 9210.79) were used to compare model fit, and produced a χ^2 statistic of 14.12 (*df* = 8, *p* = 0.08). This non-significant result suggested that Model 2 may not be an improvement in fit compared to Model 1. The coefficients were then observed for significant values to help determine the relevance of including the native language group predictors in subsequent models.

One of the four intercept coefficients was significantly different from the Spanish language group (see Table 4.8). Specifically, the Asian native language group coefficient was significant at initial status (β_{03} = 3.67, SE = 1.50, p = 0.02). However, the coefficients for the African, Austro-Asiatic, and Indo-European native language groups were not significant at initial status. LEP Status remained significant at initial status (β_{01} = -7.86, SE = 0.89, $p < 0.01$) in Model 2. None of the coded language groups showed growth rates significantly different from the Spanish language group (all $p > 0.05$). Based on the significance at initial status for one language group and for LEP Status, the Native Language variables remained in the model despite the fact that the model fit result did not strongly promote Model 2.

Table 4.8. Model 2 - Reading (Addition of Native Language Groups)

Fixed Effect	Coefficient	SE	t Ratio	df	*p*-value
Model for initial status, π_{0i}					
Initial reading score, β_{00}	345.34	0.83	417.53	367	<0.01
LEP status, β_{01}	-7.86	0.89	-8.83	367	<0.01
African language group, β_{02}	-0.60	1.48	-0.41	367	0.68
Asian language group, β_{03}	3.67	1.50	2.45	367	0.02
Austro-Asiatic language group, β_{04}	2.08	1.14	1.83	367	0.07
Indo-European language group, β_{05}	2.04	1.53	1.33	367	0.18
Model for growth rate, π_{1i}					
Initial growth rate, β_{10}	4.68	0.22	21.70	367	<0.01
LEP status, β_{11}	0.31	0.22	1.39	367	0.17
African language group, β_{12}	0.61	0.38	1.60	367	0.11
Asian language group, β_{13}	-0.06	0.27	-0.21	367	0.84
Austro-Asiatic language group, β_{14}	0.07	0.31	0.22	367	0.83
Indo-European language group, β_{15}	0.25	0.35	0.71	367	0.48

Random Effect	SD	Variance Component	df	χ^2	*p*-value
Initial status, r_{0i}	7.48	55.93	367	2807.89	<0.01
Growth rate, r_{1i}	1.67	1.36	367	580.29	<0.01
Level-1 error, e_{ti}	3.50	12.24			

The Number of Years Attending U.S. Schools predictor, a continuous variable, was added to Model 3, which still included the LEP Status and Native Language group predictors. The deviance values of Model 3 (*D* = 9191.13) and Model 2 (D = 9196.67) were used in the model comparison test. The result produced a χ^2 statistic of 5.54 ($df = 2$, $p = 0.06$), which indicated Model 3 may not be an improvement over Model 2. However, the coefficient for Number of Years Attending U.S. Schools (see Table 4.9) was significant for initial status ($\beta_{06} = -0.97$, SE = 0.45, $p = 0.03$). The slope was not significant ($\beta_{16} = -0.01$, SE = 0.11, $p = 0.93$).

Although not significant in Model 2, the Austro-Asiatic native language group coefficient was now significant at initial status ($\beta_{04} = 2.32$, SE = 1.15, $p = 0.04$) in Model 3 for reading. Continuing to demonstrate significance at initial status were the LEP Status ($\beta_{01} = -8.50$, SE = 0.94, $p < 0.01$) and Asian native language group ($\beta_{03} = 3.09$, SE = 1.48, $p = 0.04$) predictors. As observed in previous models, LEP Status and the four Native Language groups did not show growth rates significantly different from their respective reference groups. Based on the significance at initial status for Number of Years in U.S. Schools, LEP Status, and two native language groups, all variables remained in the model.

Table 4.9. Model 3 - Reading (Addition of Years in U.S. Schools)

Fixed Effect	Coefficient	SE	t Ratio	df	*p*-value
Model for initial status, π_{0i}					
Initial reading score, β_{00}	349.25	2.02	172.79	366	<0.01
LEP status, β_{01}	-8.50	0.94	-9.05	366	<0.01
African language group, β_{02}	-1.15	1.56	-0.74	366	0.46
Asian language group, β_{03}	3.09	1.48	2.08	366	0.04
Austro-Asiatic language group, β_{04}	2.32	1.15	2.02	366	0.04
Indo-European language group, β_{05}	1.92	1.50	1.29	366	0.20
Years in U.S. Schools, β_{06}	-0.97	0.45	-2.16	366	0.03
Model for growth rate, π_{1i}					
Initial growth rate, β_{10}	4.72	0.50	9.44	366	<0.01
LEP status, β_{11}	0.30	0.23	1.31	366	0.19
African language group, β_{12}	0.60	0.38	1.59	366	0.11
Asian language group, β_{13}	-0.06	0.28	-0.22	366	0.82
Austro-Asiatic language group, β_{14}	0.07	0.31	0.23	366	0.82
Indo-European language group, β_{15}	0.25	0.35	0.71	366	0.48
Years in U.S. Schools, β_{16}	-0.01	0.11	-0.09	366	0.93

Random Effect	SD	Variance Component	df	χ^2	*p*-value
Initial status, r_{0i}	7.43	55.18	366	2775.31	<0.01
Growth rate, r_{1i}	1.17	1.36	366	580.28	<0.01
Level-1 error, e_{ti}	3.50	12.24			

In the last model for consideration, the Country of Birth group predictors were added. The deviance values for this model (D = 9189.37) and Model 3 (D = 9191.13) were used to compare model fit, which produced a χ^2 statistic of 1.76 (df = 4, p > 0.50). This result suggested that Model 4 may not be a better fit compared to Model 3, and Country of Birth may not be predictive of reading achievement.

The coefficients for the two Country of Birth variables were then observed and they were not significantly different from the U.S./Canada reference group for initial status or growth (see Table 4.10). Furthermore, while LEP Status and Number of Years in U.S. Schools remained significant at initial status, the Asian and Austro-Asiatic native language predictors dropped below significance at initial status. Because model fit was not an improvement and the Country of Birth groups were not observed to be predictors of reading achievement, the Country of Birth variables were not retained in the final model.

Table 4.10. Model 4 - Reading (Addition of Country of Birth Groups)

Fixed Effect	Coefficient	SE	t Ratio	df	*p*-value
Model for initial status, π_{0i}					
Initial reading score, β_{00}	350.12	2.20	158.91	364	<0.01
LEP status, β_{01}	-8.49	0.94	-9.05	364	<0.01
African language group, β_{02}	-1.66	1.70	-0.98	364	0.33
Asian language group, β_{03}	2.52	1.66	1.52	364	0.13
Austro-Asiatic language group, β_{04}	1.88	1.24	1.52	364	0.13
Indo-European language group, β_{05}	1.53	1.62	0.94	364	0.35
Years in U.S. Schools, β_{06}	-1.09	0.47	-2.34	364	0.02
Africa/Asia/Europe group, β_{07}	0.09	1.37	0.06	364	0.95
Mexico/CA/SA/Carib group, β_{08}	-1.18	1.13	-1.05	364	0.30
Model for growth rate, π_{1i}					
Initial growth rate, β_{10}	4.77	0.56	8.57	364	<0.01
LEP status, β_{11}	0.31	0.23	1.33	364	0.19
African language group, β_{12}	0.71	0.42	1.70	364	0.09
Asian language group, β_{13}	0.02	0.34	0.07	364	0.94
Austro-Asiatic language group, β_{14}	0.14	0.34	0.41	364	0.69
Indo-European language group, β_{15}	0.34	0.38	0.90	364	0.37
Years in U.S. Schools, β_{16}	-0.02	0.12	-0.21	364	0.84
Africa/Asia/Europe group, β_{17}	-0.19	0.31	-0.62	364	0.53
Mexico/CA/SA/Carib group, β_{18}	-0.01	0.29	-0.03	364	0.98

Random Effect	SD	Variance Component	df	χ^2	*p*-value
Initial status, r_{0i}	7.41	54.98	364	2766.65	<0.01
Growth rate, r_{1i}	1.16	1.36	364	579.77	<0.01
Level-1 error, e_{ti}	3.50	12.24			

Reading – Summary

The estimated fixed effects of the final model on the effect of initial LEP status, native language groups, and number of years attending U.S. schools on reading achievement are displayed in Table 4.11. LEP Status was found to be a significant factor in reading at initial status. This result indicated that reading scores for current LEP students were 8.50 points lower than those of exited LEP students in 4th grade, producing a large effect at .92 for this reading assessment (SD = 9.28). LEP Status did not have an effect on rates of growth in reading.

Table 4.11. Final Model - Reading

Fixed Effect	Coefficient	SE	t Ratio	df	*p*-value
Model for initial status, π_{0i}					
Initial reading score, β_{00}	349.25	2.02	172.79	366	<0.01
LEP status, β_{01}	-8.50	0.94	-9.05	366	<0.01
African language group, β_{02}	-1.15	1.56	-0.74	366	0.46
Asian language group, β_{03}	3.09	1.48	2.08	366	0.04
Austro-Asiatic language group, β_{04}	2.32	1.15	2.02	366	0.04
Indo-European language group, β_{05}	1.92	1.50	1.29	366	0.20
Years in U.S. Schools, β_{06}	-0.97	0.45	-2.16	366	0.03
Model for growth rate, π_{1i}					
Initial growth rate, β_{10}	4.72	0.50	9.44	366	<0.01
LEP status, β_{11}	0.30	0.23	1.31	366	0.19
African language group, β_{12}	0.60	0.38	1.59	366	0.11
Asian language group, β_{13}	-0.06	0.28	-0.22	366	0.82
Austro-Asiatic language group, β_{14}	0.07	0.31	0.23	366	0.82
Indo-European language group, β_{15}	0.25	0.35	0.71	366	0.48
Years in U.S. Schools, β_{16}	-0.01	0.11	-0.09	366	0.93

Random Effect	SD	Variance Component	df	χ^2	*p*-value
Initial status, r_{0i}	7.43	55.18	366	2775.31	<0.01
Growth rate, r_{1i}	1.17	1.36	366	580.28	<0.01
Level-1 error, e_{ti}	3.50	12.24			

The results also showed that two native language groups, Asian and Austro-Asiatic, were strongly related to initial reading scores. These findings indicated that students in those two native language groups demonstrated significantly higher reading scores at initial status compared to students in the Spanish native language group. Again, the coefficients for the slope were not significant, which showed that native language had no effect on math growth rates when compared with the Spanish language group.

Finally, the number of years that ELL students had attended school in the U.S. had an impact on initial reading scores (β_{06} = -0.97, SE = 0.45, p = 0.03). This result indicates that, for each additional year a student had attended U.S. schools, the student earned 0.97 scale score points less in reading in 4^{th} grade.

Conclusions

The results showed that LEP Status was significantly related to initial status in both math and reading, which responds to Research Question 1. The starting point for current LEP students was 7.49 points less in math than that of exited LEP students, and 8.50 points less in reading. In both cases, the results represent nearly one standard deviation, making these meaningful effects. Many of the ELL student background variables were significantly related to initial status in math and reading, which refers to Research Question 2. Students in both the Asian and Austro-Asiatic native language groups demonstrated higher math and reading scores at initial status compared to students in the Spanish native language group.

The number of years attending U.S. schools did impact initial status in both math and reading. Country of birth had an impact on initial math scores in that students in the Mexico/Central America/South America/Caribbean country of birth group had significantly lower initial math scores than those of students in the U.S./Canada group. The analyses showed that none of the student background variables were significantly different from their respective reference groups for growth rates in either math or reading. Thus, baseline math and reading performance varied as a function of student background variables, but growth rates did not vary.

CHAPTER V

DISCUSSION

This research makes contributions to the field by offering a conceptual and methodological approach to highlighting the background characteristics of English language learners. Through the use of hierarchical linear modeling, some variation between ELL student groups was observed. By adding background variables specific to ELL students, the models became better predictors of student achievement.

The results indicated a significant relationship between LEP status and academic achievement, similar to findings from Ardasheva et al. (2012). Students classified as LEP in 4^{th} grade demonstrated lower initial mean scores in both math and reading than did exited LEP students. Students in the Asian native language group demonstrated higher math scores at initial status compared to students in the Spanish native language group, which is consistent with findings from Stiefel et al. (2003) and Roberts & Bryant (2011). Students in the Austro-Asiatic native language group also demonstrated higher math scores compared to Spanish speakers.

Students in both the Asian and Austro-Asiatic native language groups demonstrated higher reading scores at initial status compared to students in the Spanish native language group. This finding is similar to what Roberts et al. (2010) found where Asian-language ELLs had higher initial reading achievement than Spanish-speaking

ELLs. The number of years attending U.S. schools also impacted math and reading. Finally, students born in the U.S. or Canada demonstrated higher math scores in 4th grade than did students in the Mexico/Central America/South America/Caribbean country of birth group. Although this result is more geographically-specific and related to math, it is still somewhat comparable to Slama's (2012) finding where U.S.-born ELL students exhibited higher levels of academic English proficiency compared to foreign-born ELL students. The result is also related to Stevens' (1999) research where immigrants from non-Spanish speaking countries exhibited higher levels of English proficiency compared to immigrants from Spanish-language countries.

Whereas performance at initial status in math and reading varied for some groups, the growth rates did not vary. This lack of growth means that the achievement gap between some ELL student groups, like the current and exited LEP status groups, was still present after four years. This auxiliary finding reveals that students classified as LEP in 4th grade did not catch up academically to their exited LEP peers in 7th grade. This is somewhat contrary to what D'Angiulli et al. (2004) and Sung & Chang (2008) found where LEP students demonstrated higher growth in reading compared to non-LEP students. Yet, their comparisons were with LEP students and native English speakers, and this research examined the differences between current and exited LEP students. More work needs to be done in examining the long-term academic performance of current and exited LEP students.

Implications and Future Directions

These findings have implications for both future research and practice to emphasize and address the varying academic needs of ELL students. This research demonstrated that ELL students are linguistically and culturally diverse, and that these students should not be regarded as an "undifferentiated mass" (Delpit, 2006). Elaboration on this point and further implications from this study are framed in the three contexts introduced in Chapter 1: ELL Populations, Methodological Approach, and Database Management.

ELL Populations

Heterogeneity exists in the ELL student population; therefore, speaking a language other than English should not solely define them as students. As Lesaux (2006) suggested doing, this research provided additional information on the academic achievement of students formerly classified as LEP. This research also addressed the call from Genesee et al. (2005) to explore the performance of middle school students and students from different native language groups.

These findings clearly show there is diversity between native language groups, and more specifically, that linguistic diversity is present in North Carolina. The Asian native language group, including Korean or Lao speakers, demonstrated higher math and reading scores at initial status compared to the Spanish group. Vietnamese, Khmer, and Koho speakers (Austro-Asiatic group) also scored higher in math and reading compared to Spanish speakers.

Another critical element to ELL diversity is that many students in this sample are long-term ELLs. In fourth grade, 73 percent (n = 274) were classified as LEP and receiving ESL instructional services. As seventh graders, 47 percent (n = 175) of the students were still classified as LEP and receiving services in middle school. And, of this group of LEP-classified seventh graders, 87 percent (n = 153) had been in the U.S. for seven or more years, deeming them long-term English learners (Freeman & Freeman, 2002).

Many factors can contribute to this long-term status, such as below grade level proficiency in reading and writing or inconsistent schooling (Menken, Kleyn, & Chae, 2012; Olsen, 2014). The learning needs of long-term ELLs are different from newly arrived ELLs (Freeman & Freeman, 2002; Menken & Kleyn, 2009), and, if not recognized in schools and changes made in instructional programs and practices, their academic performance may not improve as a result.

Methodological Approach

As illustrated in this research, a more informative picture of ELL academic performance can be realized by examining multiple time points of test score data within a hierarchical linear model research design. The longitudinal approach used here ensured a consistent sample with stable background variables when investigating changes in mean responses and the relationship of predictors. Using HLM highlighted the variability that existed between ELL student groups and helped showcase individual predictors that affected academic achievement.

The use of HLM was advantageous in many ways. The repeated observations were not independent, but HLM handled this nesting and allowed for the study of within-student and between-student variation in math and reading achievement. The sample cohort featured differently-sized student groups, including some groups that were much larger than others, and the data were analyzed successfully. While not an issue in this study, another benefit of HLM is that it accommodates missing data which can be problematic in longitudinal studies.

Database Management

The findings in this research highlight that background differences exist in English language learners. Yet, database management has not caught up with the trends in ELL populations. For example, ELL students are not all Spanish speakers, so strategically using native language data in studies would better inform policy and practice. Recognizing such diversity has implications for improved data management and; therefore, creates great potential to fully examine ELL student performance. To start, states and individual school systems can adopt a longitudinal approach that encompasses these specific characteristics. Having access to relevant individual student information presents opportunities to tell more meaningful stories about academic performance, and perhaps lessens the need to group student data together to achieve appropriate sample sizes.

Better database management will require developing processes to collect and maintain relevant data in one central location. Typically, various departments in a school district may house their own specific data. For example, one department stores country of birth and native language information in one file and English language proficiency data in a second file, while a different department maintains end-of-year test scores. Bringing this rich information together in one central database could lead to better use of the data, and, as a result, findings can be applied toward improving instruction and academic programs.

Broadening the scope toward maintaining more ELL student-specific data is another essential area of attention. Creating and using a student information form that requests native language and country of birth is a good start toward implementing a comprehensive data management system. Collecting information on students' first language proficiency and prior schooling experiences before entering the U.S. would also be informative and useful for schools and as areas of study. Additionally, English language proficiency data should be stored in a more complete manner, such as by including the scores, test name, and the domain tested. Maintaining proficiency data in this manner could help eliminate inconsistencies and add more clarity to future research.

Similar to inconsistencies in how ELL, LEP, and ESL terms are defined and used in research (Abedi et al., 2000; deJong, 2004; Maerten-Rivera et al., 2010), these problems persist when tracking ELL performance over time. When students exit LEP status, they no longer attend ESL instruction. Students who exit LEP status are still considered members of the LEP subgroup for two years for accountability purposes. However, when those two years end, their test score data are no longer linked to the LEP subgroup. As a result of this policy, exited LEP students are often erroneously considered as never-ELL students, and, so, academic progress has been difficult to track and assess.

Improvements in database management must be instituted to capture and follow long-term performance more consistently. One change is to code student data more accurately by clearly specifying in a separate field that the student is an ELL, in addition to noting the number of years participating in ESL instruction and the LEP exit year in other fields. Another change would be to incorporate Olsen & Jaramillo's three categories for older English learners (Freeman & Freeman, 2002) into the database structure. This revision would help in identifying ELL students based on their prior schooling experiences and arrival in the U.S., which in turn can bring to light the long-term English learners who are still in need of support.

By not maintaining such a comprehensive database, the ability to perform sophisticated work is lost and the wealth of information about the diversity of ELL students is squandered. Basic and applied research that could be conducted does not occur, and our understanding of ELL academic performance is diminished because of inadequate data management.

Limitations and Future Research

The intent of this research was to demonstrate the heterogeneity of English language learners, yet there were still limitations. Although the data were collected from a school district with over 60,000 students, the final sample size was 373 students using the four-year cohort design. Having access to individual student background information was instrumental in going forward with this research. Nevertheless, the subsequent sizes of the student variable groups still presented challenges.

There were not enough data to recognize multiple native languages. As a result, the students' 36 languages were placed into five groups using a framework with Spanish representing one of those groups. Similar to sample groups used in previous research studies (Hofstetter, 2003; Kieffer, 2008; Robinson, 2008; Hopstock & Stephenson, 2003), the Spanish native language group was considerably larger than the other four language groups. Another limitation was that my framework design may have specifically contributed to this study's findings. Other schemas could have been followed to organize the data into meaningful groups, such as considering language difficulty or similarity to English. Using a different schema to group the native languages might have produced different results.

Strategic grouping was also necessary for the country of birth variables. The heterogeneity of the sample was evident in the students' 35 different birth countries, although 43 percent of students in the sample were born in the United States and Canada and 20 percent were born in Mexico. If the data had permitted a more extensive analysis of native languages or countries of origin, additional heterogeneity may have been revealed. More informative results might have then allowed additional questions to be addressed, such as whether reading comprehension skills are more difficult to master for some students based on their native language.

Future research may examine the interaction effects within different native languages and birth countries. Examining potential interactions would allow for deeper investigation into ELL student diversity. Also, studying other variables, such as prior schooling experiences or student perceptions and attitudes, in combination with native languages could further highlight the heterogeneity within ELL groups. A three-level model could be used to look at variability among schools in terms of socioeconomic status or school climate. Another potential area for future research is to study possible confounding issues in the data. Examples of such confounds are number of years in U.S. schools with native language or country of birth with native language. These in-depth research opportunities will be dependent upon sufficient sample sizes available for use.

Another consideration is for states and school districts to make better use of the data they own, which would begin with improved data collection methods. Larger districts can dig deeper into their ELL student performance by collecting and then incorporating relevant background characteristics into evaluations and longitudinal studies. A good start would be to access and use variables similar to those used in this study, such as creating relevant native language groups that would highlight as many languages as possible. Examining LEP status as a time-varying predictor, number of years classified as LEP, graduation rates, or ESL program type are also meaningful opportunities for learning about ELL achievement. The efforts to track the long-term performance of ELL students have been limited, and this study just touches the issue. Yet with better data management, there will be more opportunity to examine that performance. Taking a strong look at all available data could produce results that would be helpful in informing instruction and improving programs for ELL students.

REFERENCES

Abedi, J. (2002). Standardized Achievement Tests and English Language Learners: Psychometrics Issues. *Educational Assessment*. 8 (3), 231-257.

Abedi, J. (2004). The No Child Left Behind Act and English Language Learners: Assessment and Accountability Issues. *Educational Researcher*. 33 (1), 4-14.

Abedi, J. (2008). Classification System for English Language Learners: Issues and Recommendations. *Educational Measurement: Issues and Practice*. 27 (3), 17-31.

Abedi, J. & Lord, C. (2001). The Language Factor in Mathematics Tests. *Applied Measurement in Education*. 14 (3), 219-234. Los Angeles: University of California, Center for the Study of Evaluation and National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of Accommodation Strategies on English Language Learners' Test Performance, *Educational Measurement: Issues and Practice*. 19 (3), 16-26.

Ardasheva, Y., Tretter, T.R., & Kinny, M. (2012). English Language Learners and Academic Achievement: Revisiting the Threshold Hypothesis. *Language Learning*. 62 (3), 769-812.

Aristar, A.R. (1990). *The Linguist List*. Institute for Language Information and Technology, Eastern Michigan University. www.linguistlist.org.

Bankston, C.L. & Zhou, M. (1995). Effects of Minority-Language Literacy on the Academic Achievement of Vietnamese Youth in New Orleans. *Sociology of Education*. 68, 1-17.

Bryk, A.S. & Raudenbush, S.W. (1987). Application of Hierarchical Linear Models to Assessing Change. *Psychological Bulletin*. 101 (1), 147-158.

Callahan, R.M. (2005). Tracking and High School English Learners: Limiting Opportunity to Learn. *American Educational Research Journal*. 42 (2), 305-328.

Chalhoub-Deville, M. & Deville, C. (2008). Nationally Mandated Testing for Accountability: English Language Learners in the US. In Spolsky, D. & Hunt, F.M. (Eds.), The Handbook of Educational Linguistics. Malden, MA: Blackwell Publishing.

Chang, M. (2008). Teacher Instructional Practices and Language Minority Students: A Longitudinal Model. *The Journal of Educational Research*. 102 (2), 83-97.

Chen, F., (2010). *Differential language influence on math achievement*. Retrieved from NC Digital Online Collection of Knowledge and Scholarship (NCDOCKS). Greensboro, NC: University of North Carolina at Greensboro.

Chen, F. & Chalhoub-Deville, M. (2014). Principles of quantile regression and an application. *Language Testing*. 31 (1), 63-87.

Cummins, J. (2009). Transformative multiliteracies pedagogy: Schoolbased strategies for closing the achievement gap. *Multiple Voices for Ethnically Diverse Exceptional Learners*. 11.2 p. 38-56. Division for Cultural & Linguistically Diverse Exceptional Learners for the Council of Exceptional Children

D'Angiulli, A., Siegel, L.S., & Maggi, S. (2004). Literacy Instruction, SES, and Word-Reading Achievement in English-Language Learners and Children with English as a First Language: A Longitudinal Study. *Learning Disabilities Research & Practice*. 19 (4), 202-213.

deJong, E.J. (2004). After Exit: Academic Achievement Patterns of Former English Language Learners. *Education Policy Analysis Archives*. 12 (50), 1-18.

Delpit, L. (2006). Other People's Children: Cultural Conflict in the Classroom. New York: The New Press.

Deyhle, D. (2009). Reflections in Place: Connected Lives of Navajo Women. Tucson: The University of Arizona Press.

Durán, R.P. (2008). Assessing English-Language Learners' Achievement. *Review of Research in Education*. 32, 292-327.

Freeman, Y.S. & Freeman, D.E. (2002). Closing the Achievement Gap: How to Reach Limited-Formal-Schooling and Long-Term English Learners. Portsmouth, NH: Heinemann

Garcia, G. E. (2000). Bilingual children's reading. In M. Kamil, P. Mosenthal, D. Pearson, R. Barr (Eds.), Handbook of Reading Research, Volume III. Newark, DE: International Reading Association.

Genesee, F., Lindholm-Leary, K., Saunders, W., & Christian, D. (2005). English Language Learners in U.S. Schools: An Overview of Research Findings. *Journal of Education for Students Placed at Risk*. 10 (4), 363-285.

Goh, D.S. (2004). Assessment Accommodations For Diverse Learners. Boston, MA: Pearson.

Hofstetter, C.H. (2003). Contextual and Mathematics Accommodation Test Effects for English-Language Learners. *Applied Measurement in Education*. 16 (2), 159-188.

Hopstock, P.J. & Stephenson, T.G. (2003). *Descriptive Study of Services to LEP Students and LEP Students with Disabilities: Special Topic Report #1 Native Languages of LEP Students*. U.S. Department of Education, Office of English Language Acquisition, Language Enhancement, and Academic Achievement for Limited English Proficient Students (OELA).

Kieffer, M.J. (2008). Catching Up or Falling Behind? Initial English Proficiency, Concentrated Poverty, and the Reading Growth of Language Minority Learners in the United States. *Journal of Educational Psychology*. 100 (4), 851-868.

Kim, J. & Herman, J.L. (2009). A Three-State Study of English Learner Progress. *Educational Assessment*. 14, 212-231.

LaCelle-Peterson, M.W. & Rivera, C. (1994). Is It Real for All Kids? A Framework for Equitable Assessment Policies for English Language Learners. *Harvard Educational Review*. 64 (1), 55-76.

Lee, M. & Madyun, N. (2008). School racial composition and academic achievement: the case of Hmong LEP students in the USA. *Educational Studies*. 34 (4), 319-331.

Lesaux, N.K. (2006). Building Consensus: Future Directions for Research on English Language Learners at Risk for Learning Difficulties. *Teachers College Record*, 108 (11), 2406-2438.

Lesaux, N.K., Rupp, A.A., & Siegal, L.S. (2007). Growth in Reading Skills of Children from Diverse Linguistic Backgrounds: Findings from a 5-Year Longitudinal Study. *Journal of Educational Psychology*. 99 (4), 821-834.

Lewis, M. Paul (Ed.), 2009. Ethnologue: Languages of the World, Sixteenth Edition. Dallas, TX: SIL International. Online version: http://www.ethnologue.com/

McBrien, J.L. (2005). Educational Needs and Barriers for Refugee Students in the United States: A Review of the Literature. *Review of Educational Research*, 75 (3), 329-364.

Macedo, D. (2006). Literacies of Power: What Americans Are Not Allowed to Know. Boulder: Westview Press.

Maerten-Rivera, J., Myers, N, Lee, O., & Penfield, R. (2010). Student and School Predictors of High-Stakes Assessment in Science. *Science Education*. 94 (6), 937-962.

Menken, K. & Kleyn, T. (2009). The Difficult Road for Long-Term English Learners. *Educational Leadership*. 66 (7).

Menken, K., Kleyn, T., & Chae, N. (2012). Spotlight on "Long-Term English Language Learners": Characteristics and Prior Schooling Experiences of an Invisible Populations. *International Multilingual Research Journal*. 6, 121-142.

North Carolina Department of Public Instruction (2014). *Guidelines for Testing Students Identified as Limited English Proficient*. Academic Services and Instructional Support. Division of Accountability Services/North Carolina Testing Program. Retrieved from http://www.ncpublicschools.org/docs/accountability/policyoperations/lep/testlepstudents14.pdf

Olsen, L. (2010). <u>Reparable Harm: Fulfilling the unkept promise of educational opportunity for Long Term English Learners</u>. Oakland, CA: California Tomorrow and Los Angeles County Office of Education.

Olsen, L. (2014). <u>Meeting the Unique Needs of Long Term English Language Learners.</u> National Education Association.

Purcell-Gates, V. (1995). <u>Other people's words: the cycle of low literacy</u>. Cambridge, MA: First Harvard University Press.

Raudenbush, S.W., Bryk, A.S., Cheong, Y.F., Congdon Jr., R.T., & duToit, M. (2011). <u>HLM 7: Linear and Nonlinear Modeling</u>. Lincolnwood, IL: Scientific Software International, Inc.

Raudenbush, S.W. & Bryk, A.S. (2002). <u>Hierarchical Linear Models: Applications and Data Analysis Methods</u>. Thousand Oaks, CA: Sage Publications.

Roberts, G. & Bryant, D. (2011). Early Mathematics Achievement Trajectories: English-Language Learner and Native English-Speaker Estimates, Using the Early Childhood Longitudinal Survey. *Developmental Psychology*. 47 (4), 916-930

Roberts, G., Mohammed, S.S., & Vaughn, S. (2010). Reading Achievement Across Three Language Groups: Growth Estimates for Overall Reading and Reading Subskills Obtained With the Early Childhood Longitudinal Study. *Journal of Educational Psychology*. 102 (3), 668-686.

Robinson, J.P. (2008). Evidence of a Differential Effect of Ability Grouping on the Reading Achievement Growth of Language-Minority Hispanics. *Educational Evaluation and Policy Analysis*. 30 (2), 141-180.

San Miguel, G. & Donato, R. (2009). Latino Education in Twentieth-Century America: A Brief History. In Murillo Jr., E.G., Villenas, S., Trinidad-Galvan, R., Munoz, J.s., Martinez, C., & Machado-Casas, M. (Eds.), Handbook of Latinos and Education: Theory, Research, and Practice. New York: Routledge.

Saavedra, C.M., Morrison, S., Smith, P.H., & Bailey, Raleigh (2008). *Immigrant Families in Guilford County: A Report*. Retrieved from University of North Carolina, Greensboro, The Center for New North Carolinians website: http://cnnc.uncg.edu/wp-content/uploads/2012/12/Immigrant-Families-in-Guilford-County.pdf

Schoorman, D. (2001). Addressing the Academic Needs of Immigrant Students: Issues and Trends in Immigrant Education. In Diaz, C.F., Multicultural education in the 21st Century. New York: Addison-Wesley.

Short, D. & Fitzsimmons, S. (2007). *Double the work: Challenges and solutions to acquiring language and academic literacy for adolescent English language learners - A report to Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education.

Slama, R.B. (2012). A Longitudinal Analysis of Academic English Proficiency Outcomes for Adolescent English Language Learners in the United States. *Journal of Educational Psychology*. 104 (2), 265-285.

Snijders, T.A.B. & Bosker, R.J. (1999). *Multilevel Analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Publications.

Stevens, G. (1999). Age at Immigration and Second Language Proficiency among Foreign-Born Adults. *Language in Society*. 28 (4), 555-578.

Stiefel, L., Schwartz, A.E., & Conger, D. (2003). Language Proficiency and Home Languages of Students in New York City Elementary and Middle Schools. New York: Taub Urban Research Center, New York University.

Sung, Y.Y. & Chang, M. (2008). Center-based care for language minority students. *Educational Research and Evaluation*. 14 (5), 445-463.

Wang, J. & Goldschmidt, P. (1999). Opportunity to Learn, Language Proficiency, and Immigrant Status Effects on Mathematics Achievement. *The Journal of Educational Research*. 93 (2), 101-111.

WIDA (2010). *The Cornerstone of WIDA's Standards: Guiding Principles of Language Development*. https://www.wida.us/aboutus/AcademicLanguage/.

Wolf, M.K., Herman, J.L. & Dietel, R. (2010, Spring). *Improving the Validity of English Language Learner Assessment Systems* (CRESST Policy Brief No. 10 - Full Report). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

APPENDIX A

ACRONYMS AND DEFINITIONS

ELL - An English language learner is a national-origin-minority student who is limited English proficient. This term is often preferred over limited English proficient (LEP) as it highlights accomplishments rather than deficits *

ESL - English as a Second Language is a program of techniques, methodology and special curriculum designed to teach ELL students English language skills, which may include listening, speaking, reading, writing, study skills, content vocabulary, and cultural orientation. ESL instruction is usually in English with little use of native language *

FEP - Fluent (or fully) English proficient *

IFE - Students who are initially fluent in English demonstrate sufficient initial fluency in English as to not require specialized support services.

LEP - Limited English proficient *

SES - Socioeconomic status; often determined as "low" by family qualification for Free/Reduced Lunch status

Redesignated - Students who were former LEP students and redesignated as FEP (fluent English proficient), according to the multiple criteria, standards, and procedures adopted by the respective district or state.

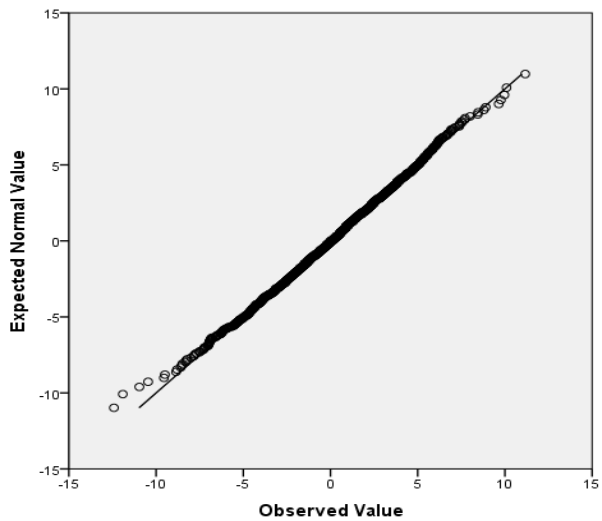
Language proficiency - Refers to the degree to which the student exhibits control over the use of language, including the measurement of expressive and receptive language skills in the areas of phonology, syntax, vocabulary, and semantics and including the areas of pragmatics or language use within various domains or social circumstances. Proficiency in a language is judged independently and does not imply a lack of proficiency in another language *

* http://www2.ed.gov/about/offices/list/ocr/ell/glossary.html

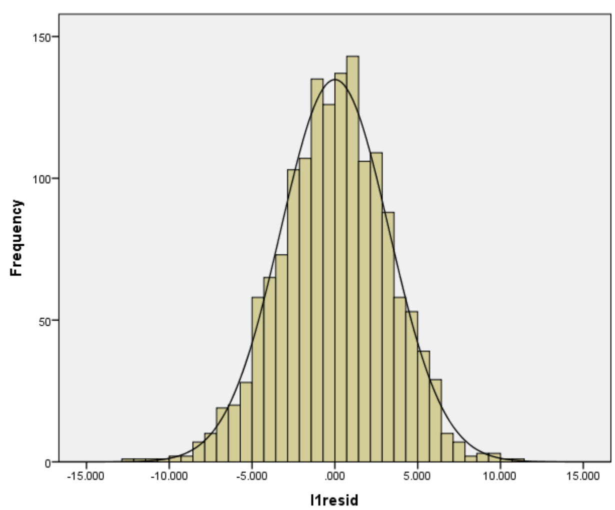
APPENDIX B

MATH RESIDUALS

Q-Q plot of Level-1 Math Residuals



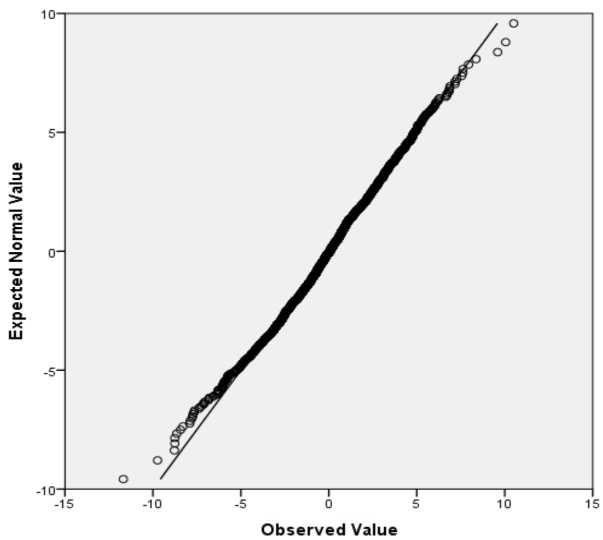
Histogram of Level-1 Math Residuals



APPENDIX C

READING RESIDUALS

Q-Q plot of Level-1 Reading Residuals



Histogram of Level-1 Reading Residuals