UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

HOLDING BACK ENGLISH LEARNERS: THE IMPACT OF EARLY ELEMENTARY GRADE RETENTION ON LANGUAGE DEVELOPMENT

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the

Degree of

DOCTOR OF EDUCATION

By

JAMIE LAKE BUCKMASTER

Norman, Oklahoma

2019

HOLDING BACK ENGLISH LEARNERS: THE IMPACT OF EARLY ELEMENTARY GRADE RETENTION ON LANGUAGE DEVELOPMENT

A DISSERTATION APPROVED FOR THE

DEPARTMENT OF EDUCATIONAL LEADERSHIP AND POLICY STUDIES

BY

Dr. Angela Urick, Chair

Dr. Lawrence Baines

Dr. Tim Ford

Dr. Jeffrey Maiden

Dr. Courtney Vaughn

Abstract

An abundance of scholarly work indicates that the practice of requiring a student to repeat a grade likely has significant academic, social and long-term effects on students. Grade retention, prevalent in elementary schools across the United States, is even more common for English learners, who make up more than 10% of students across the country, and who are already more likely to struggle in school. It is difficult to find literature to inform practitioners and scholars of the outcomes of EL students' language development after grade retention. The present study seeks to examine whether grade retention affects EL students, and whether there may be detriments to language development.

The past decades have yielded many studies on the topic of grade retention, with findings indicating negative outcomes for students in the years after they repeat a grade, such as higher drop-out rates, lower academic achievement, and social and emotional harm. Despite the literature, the practice of requiring a student to remain in the same grade for more than one year has grown to be a common practice. This is particularly alarming for English learners, who have a statistically higher likelihood of being retained. At the same time, English learners arrive at school with a number of risk factors, including higher risk of dropping out of high school and a greater likelihood of feeling disenfranchised and unwelcome in schools.

English learners are more likely than English-proficient peers to be retained in grade. Perhaps this is no surprise, since EL students are, by definition, in the process of learning English, and are often perceived to be behind academically when they cannot demonstrate their knowledge in English. This, in combination with a lack of proficiency in reading, perhaps leads educators to retain EL students much more often than their classmates.

A weighty gap exists in the current literature concerning the crossroads of these issues. The present study seeks to inform scholars and practitioners of the outcomes of retention for English learners by analyzing the language development over time for EL students seven years after retention, as compared with promoted EL peers of statistical similarity pre-retention. The sample included students who were current English learners in a district in the Midwest. Propensity score matching was utilized to create quasi-experimental comparison groups of retained, treatment, and matched-promoted students, control. Data were analyzed using a multilevel model to examine change between individuals and over time. As expected, on the whole, students in both groups improved their English proficiency over time; however, when compared with matched-promoted peers, retained English learners achieved lower scale scores in all three assessed domains: Listening, Reading and Writing. The growth slopes did not significantly differ between retained and promoted groups for any domain.

When English learners are held back, their language achievement is lower. Students can never meet the linguistic performance of their same-aged peers. By holding back English learners, teachers sort students into a lower grade rather than intervening to increase their growth so that they can catch up to perform at grade level.

Acknowledgements

"If I have seen a little further it is by standing on the shoulders of Giants."

– Isaac Newton

Some have described the dissertation process as a journey; I prefer to liken it to a roller coaster with fire obstacles. I am extremely fortunate to have had the mentorship of Dr. Angela Urick to guide me through this process. Dr. Urick, thank you for the time, effort and thought you have put into me throughout this project. I'm quite sure that you hesitated the first time that I asked you what propensity score matching was, just like you sighed when I wanted to include all three dependent variables, but you let me take on challenges (read: bite off more than I can chew), bailing me out as needed, and I'm so much better for it. Thank you for your patience, and for helping me grow.

I'd like to extend my deepest gratitude to my committee members: Dr. Courtney Vaughn, (who taught me to code and theme with the best of them), Dr. Jeffrey Maiden, Dr. Tim Ford, and Dr. Lawrence Baines. I am also grateful for the wisdom and direction provided by Dr. Bill Frick. You are all giants, and I'm thankful to have learned from you.

Dedication

I am a product of my family, and I am grateful to those who have shaped me.

To my mom, the strongest, fiercest, most amazing woman I know,
who taught me the joy of being an educator and instilled a deep fear of grammar mistakes,
I aspire daily to have a heart as big as yours.
To my dad, who has always known everything, even when he was making it up,
and who used to write a limerick every day for my lunch box,
you taught me the power of both writing and humor, two of my best tools,
to Steve, who showed me that putting one foot in front of the other will get you there,
to my Granddad, the first Sooner in our family,
who showed me the joy of working hard and loving what you do,
to my grandparents together, who showed us what family looks like,
and to my husband and best friend, my fiercest competition and closest ally,
you are my constant and my rock.

Thank you.

Table of Contents

Chapter 1: Introduction 1
Description of the Study 2
Purpose 2
Significance 3
Background to the Study 7
Educating English learners in the United States 8
A History of Grade Retention 14
Post-Civil War to Early 1900s 14
Early retention research 15
A Nation at Risk 16
A push for accountability 16
Conclusion 18

Chapter 2: Literature Review 20
A Clustering of Disparities 20
Grade Retention and Academic Achievement 24
Retention and English learners: A Clustering of Disparities 34
Framework for This Study 37
Research Question 38

Chapter 3: Methods 40
Introduction 40
Setting 40
Table 1. English Learners in OKCPS 41
Table 2. Student Retentions by Year 42
Instrumentation 43
Procedure 45
Initial Data Tests 47
Propensity Score Matching 50
PSM Model Design 51
Covariates for Matching 51

Specifications for Determining Propensity Score 56
Data Analysis 58
Measuring the Research Question: Linguistic Growth 58
Data Restructure and Cleaning 58

Chapter 4: Results 61
Propensity Score Matching 61
Propensity Score Models 61
Longitudinal Analysis 67

Chapter 5 79
Introduction 79
Study Findings, Past and Future Research 80
Implications for Policy and Practice 86
Limitations 89
Future Research 89
Conclusion 90

References 92

Appendix A. Repeated Measures ANOVA 117
Appendix B. Propensity Score Matching 121
Appendix C. Linear Trends 125
Appendix D: Input Instructions for MPlus 129
Appendix E. Results – Unweighted vs Weighted 132

Chapter 1: Introduction

English learners (ELs), a growing subgroup of students in US schools (NCES, 2019), are more likely than peers to be retained in grade (Tingle, Schoeneberger & Algozzine, 2012; Winsler, Hutchison & De Feyter, 2012; Willson & Hughes, 2006; Tillman, Guo & Harris, 2006). However, an abundance of scholarly literature indicates that retention is unhelpful and possibly even harmful for students (Klapproth, Schaltz, Brunner, Keller, Fischbach, Ugen & Martin, 2016; Vandecandelaere, Vansteelandt, De Fraine, & Van Damme, 2015; Im, Hughes & Kwok 2013). Further, the practice of retaining EL students specifically may be more even more detrimental, as such students already struggle socially and academically in schools (Lamote, Pinxten, Van Den Noortgate, & Van Damme, 2014; Sox, 2009; Wu, West & Hughes, 2008).

Grade retention has been controversial in public schools for decades. Many teachers and administrators hold fervently to beliefs for or against the practice of grade retention. The issue has become a debate in Oklahoma, where state law mandates retention in third grade for students who are not reading on grade level. For Oklahoma City Public Schools, the state law has had the additional effect of increasing the numbers of non-mandatory retentions. This has been shown in other regions as well (Goos, Van Damme, Onghena, Perty, & de Bilde, 2013; Schwager, Mitchell, Mitchell, & Hecht, 1992) where, as state-mandated retentions have increased, so have non-mandatory, teacher-initiated retentions. This happens as grade retention becomes the norm: a go-to solution for educators in a climate where retention is endorsed by the state or country (Goos, Van Damme, Onghena, Perty & de Bilde, 2013). Quite often, retention decisions are based upon teacher opinion rather than data and evidence (Beebe-Frankenberger, Bocian, MacMillan & Gresham, 2004). Because retention has often been shown to have ill social and academic effects (Martin, 2011; Chen, Liu, Zhang, Shi, & Rozelle, 2010) such decisions cannot

be made hastily. Still, no studies have undertaken this question as it pertains to retention of EL students, who are identified for grade retention more often than English-only peers (Tingle, Schienverger & Algozzine, 2012; Winsler et al., 2012).

Description of the Study

The present study will investigate the language development outcomes of EL students who have been retained in grade with a quasi-experimental, longitudinal design. Using propensity score matching, groups of retained and matched-promoted students will be assessed throughout a seven-year period to gauge their language outcomes. Students will be matched on background and academic variables. An analysis of the data, including a two-level model for individual change, will indicate whether there are differences in language achievement and growth among retained and promoted students. Results will shed light on the problem of retention as it relates to English learner students.

Purpose

This study will add to the growing body of research on the common practice of retaining students. Though much evidence exists regarding adult outcomes and socio-emotional outcomes for retained students, a gap exists in the literature concerning outcomes for English learner students, who are distinct from their peers. Empirical evidence is needed to confirm whether literature surrounding the academic outcomes of retained students can be applied to English learners. Therefore, the purpose of this study is to demonstrate the impact of grade retention on language outcomes for ELs compared to their non-retained, matched peers. Based on previous literature, the present study offers the hypothesis that grade retention has a negative or negligible effect on EL students' language development.

Significance

The present study contributes to a broader discussion about how a vulnerable population, English learner students, may be negatively affected by grade retention, both because they are more likely to be identified for retention, and because, as the current study indicates, their language development may be negatively impacted in significant ways.

The most recent census data indicate that the United States is experiencing a historically high number of immigrants, with over 44 million individuals in the country (Zong, Batalova & Burrows, 2019). Of these, roughly 48% of individuals over five years old rated their English as not proficient (Zong, Batalova & Burrows, 2019). In other words, many residents of the United States are not yet proficient in a language that they need in order to fully access their surroundings. This is especially true for students, who must learn and show achievement in English in order to be considered successful in school.

English learners arrive in schools with several obstacles to overcome. Students need English proficiency in order to comprehend and show understanding in school. Though it takes just six months to two years to attain social language, students need five to seven years to attain proficiency in advanced academic language that they need in order to succeed in school (Cummins, 1981). Proficiency can take up to ten years if students are not literate in their first language (Cummins, 1981). As such, students must simultaneously learn academic language and content for roughly half of their years in school.

Around 23% of students in American schools are the children of immigrants (Crosnoe & Turley, 2011). The process of immigration itself, coupled with the surrounding difficulties, is important to consider. Many families move to the United States to seek a better life for their children. Such a move often requires leaving behind family, friends and support systems, only to

be confronted with a hard life, too many hours of work, and strained familial relationships due to stress in the first years in their new country (Bacallao & Smokowski, 2007).

Students often leave behind the difficult or hazardous conditions in which they have grown up. Some students arrive as refugees, having fled their home countries with little to no prior schooling; this is particularly true of students arriving from Latin America (Crosnoe & Turley, 2011). Though each country's education system differs and evolves over time, Schaller, Rocha, and Barshinger (2007) found that 91% of Mexico's school-age population in the 1970s did not complete sixth grade. Today, half of Mexican immigrant children have no parent with a high school degree, which Crosnoe and Turley characterize as a "clustering of disparities" (2011, p. 145): students have left education systems, some of which may be very poor quality, to arrive in the United States with both a language and content gap, compounded with a lack of academic support from parents, who, though eager to help students, are unable to provide assistance such as tutoring and homework aid.

Many English learner students also come from mixed-status homes, meaning that one parent is of undocumented status. Recent census data estimates that approximately 4 million US-born children under the age of 18 have at least one undocumented parent (Gelatt & Zong, 2018). Approximately one in ten families with children is of mixed status, with significantly higher numbers in large cities such as Los Angeles and New York City, where up to 40% of low-income families are of mixed status (Fix & Zimmerman, 2001). Though 89% of the children in mixed-status homes are US citizens, they are much less likely to have access to services such as food stamps, health care, and early childhood education due to a misunderstanding or distrust in government aid (Yoshikawa & Kalil, 2011). In addition to the stress of poverty is an underlying

fear and dread of a parent being deported, causing a constant undercurrent of fear of the family being split.

EL students arrive in US schools to confront the added challenge of learning both language and content simultaneously. Though some countries are more advanced in math and science than equivalent curricula across the United States, demonstrating knowledge of such concepts in English is challenging. Some subjects, such as US history, may be entirely new for students. Still more students arrive with large gaps in education due to limited access to schools in their home countries.

Students arriving as teenagers immediately face the challenge of meeting high school graduation requirements as they must simultaneously navigate a new, unfamiliar school system and learn in a new language. High school exit exams require students to demonstrate success in English on standardized exams within four years of learning English. A requirement to attain enough credits to graduate is often at odds with high rates of absenteeism as students fulfill roles in helping parents or family members during the day, such serving as an interpreter for a parent's appointment at the doctor, or because the students themselves work jobs to help support their families. Perhaps, then, it is no wonder that EL students have the highest rates of dropping out of high school. In the 2013-14 school year, the Average Cohort Graduation Rate (ACGR) for English learners was 63 percent, as compared to the overall national ACGR of 82 percent (NCES, 2018). English learners fell behind non-EL peers in every single state (NCES, 2018).

In Oklahoma, English learners are seven percentage points behind the average for high school completion rates (McFarland, Cui & Stark, 2018). Moreover, a new report has shown that Oklahoma is second in the nation for grade retention rates (Palmer, 2018).

With more EL students than any other district in the state, combined with a new state law mandating retention for students not passing the third grade reading test, Oklahoma City Public Schools is in dire need of a more information about how grade retention may affect this large subgroup. English learners are more likely to be retained than English-proficient peers nationwide (Tingle, Schoeneberger & Algozzine, 2012). This is also true for OKCPS: internal data analysis has shown that EL students in Oklahoma City Public Schools are two to three times more likely to be retained when compared with English-only peers. Teachers express support for retaining EL students in particular based on allowing them more time for language growth, despite the fact that best practices promote simultaneous language and content development. Because teachers cite academics as a primary reason for retaining students, the present study seeks answers about the academically-dependent language outcomes of English learner students after retention in order to provide empirical evidence for teachers and administrators both within Oklahoma City Public Schools, and in other districts serving EL students.

The outcomes of grade retention have been investigated thoroughly in the realms of academic, social and long-term outcomes for all students; however, no studies have specifically set out to investigate language trajectories for English learner students over time. English learner students are a fast-growing sub-population of schools in the United States, growing from 8.7 percent of the school population in 2002-03 to 9.1 percent in 2011-12 (NCES, 2014). If initial investigations indicate that EL students may be harmed by retention more than English-only students, and the literature indicates that EL students already feel unwelcome (Sox, 2009) and are more likely to drop out (Sheng, 2011), then such investigations are imperative and overdue.

Since No Child Left Behind, the beginning of the accountability movement, and a call to increase rigor and standards for all children, a growing number of states have adopted legislation

similar to Oklahoma's Reading Sufficiency Act (Huddleston, 2015), which mandate grade retention, sometimes combined with additional interventions, for students that have failed state reading tests. Though the Reading Sufficiency Act makes allowances for English learner students, it only grants reprieve from retention for EL students who have been in the country for fewer than two years. Cummins (1981) hypothesized that students need five to seven years to acquire academic language before they are proficient in English. Cook, Boals and Lundberg (2011) indicated that this is consistent with nationwide English language proficiency data. In fact, Krashen (2015) explains that English learner students are, by definition, not proficient enough to pass standardized exams. Thus, theory, research and best practices as they relate to English learner students and to grade retention in general are in conflict with such policies. However, more answers are needed regarding the intersection of these two issues, informing educators of the outcomes of grade retention specifically for English learners.

Background to the Study

The landscape of education has changed dramatically over the decades, from one-room schoolhouses to high-stakes testing, accountability, and frequent public outrage at the state of schools (Hunt, 2008). One consistency has been the immigration of millions of individuals to the United States, following their dreams and seeking refuge in a new place (Crosnoe & Turley, 2011). Throughout the history of the country, the children of immigrants have been shuffled into schools to receive education in American schools (Wright, 2005). The education of immigrant students has evolved and expanded; it has changed consistently through time, with student and family rights becoming more defined each decade. Schools receiving federal monies are now

required to serve English learner (EL) students in programs that allow them equitable access so they may be afforded the same opportunities as their peers (Wright, 2005).

Unlike the evolving rights of and programming for English learners, few issues have cropped up over this span of time in education, weathering changes to remain an unresolved, controversial modern dilemma. From the beginning, the question of how and when to promote or retain students in the same grade has been volleyed up, disputed, settled into habit or policy, and then soon discovered anew, triggering the cycle again (Lindvig, 1983; Abidin, Golladay & Howerton, 1972; Coffield & Blommers, 1956; Cook, 1941; Keyes, 1911).

Today, the topics of programming for English learners and grade retention converge into a bigger, more complex issue facing educators: grade retention as a benefit or detriment to EL students. This section seeks to follow both journeys historically, intertwining and concatenating over time, and then to delve into modern issues and questions relating to this topic.

Educating English learners in the United States

The question of how to serve students with a native language other than English is undoubtedly as old as the first day that a non-English speaker arrived for his or her first day of school; however, the policies and requirements of schools are easier to trace through decisions from the Supreme Court, starting with *Plessy v. Ferguson*, allowing "separate but equal" facilities for students (*Plessy v. Ferguson*, 1896). Though this is widely considered to have targeted African Americans, various minority groups were also separated based on language proficiency (Greenfield & Kates, 1975).

Plessy and Brown. Three cases were considered by higher courts between the *Plessy* and *Brown v. Board of Education (1954)* rulings. These cases shed light on the struggles of English learner students and their families during this time. In *Del Rio Independent School District v.*

Salvatierra (1930), parents and families challenged their school district's policy of serving children with a Latino surname in different schools. Texas courts sided with the schools that the segregation was necessary and permissible for teaching English. A similar case, *Mendez v. Westminster School District (1947),* a trial court in California ruled that separate schools were not adequately meeting the needs of students, compelling school districts to serve students together under the law. Another case in Texas, *Delgado v. Bastrop Independent School District (1948)* prohibited the practice of serving English learner students in separate facilities due to language proficiency. However, despite such rulings, de facto segregation continued (Contreras & Valverde, 1994). The United States Supreme Court, in *Brown v. Board of Education (1954)*, banned "separate but equal" as a practice, prohibiting racial separation in schools; however, Hispanic students were not recognized as a separate race at the time (Contreras & Valverde, 1994). It was not until *Cisneros v. Corpus Christi ISD (1970)* that the courts identified Mexican-Americans to be a separate ethnic group against whom school districts could not discriminate.

The Civil Rights Act. The Civil Rights Act of 1964 was a major advance for the rights of language minority students in schools by way of the Equal Protection Clause, which banned discrimination based on race and national origin. National origin would come to include English learner students and provide protections in schools receiving federal monies. The Office of Civil Rights further clarified the rights of English learner students on a memo dated May 25, 1970, indicating that districts should take steps to assist students with language difficulties. What exactly schools and districts were required to provide was often contested until a group of parents of Chinese American families in California sued the San Francisco Public Schools Board of Education, winning a landmark case, *Lau v. Nichols (1964)*. This decision required school districts to identify English learner students, evaluate their language skills, provide interventions

to address language difficulties, include EL students in mainstream classes, and require teachers to address language gaps of students. *Lau* effectively set up the foundation for all English language development programming for public schools, mandating that students with a native language other than English must be included and brought up to the level of their English-only peers.

Castañeda. The next decade, additional clarification was brought about in another Supreme Court case. *Castañeda v. Pickard* went beyond the nebulous requirements of *Lau* to require districts to have English language development (ELD) programs that meet three standards. The *Castañeda* Standard requires schools' ELD programs to be based on sound educational theory, implemented effectively via both personnel and resources, and evaluated for effectiveness. From this point, schools were not only responsible for addressing English proficiency, but for ensuring that students were closing the achievement gap with English-only peers.

Unfortunately, even though rulings such as *Lau* and *Castañeda* required schools to address language gaps of students enrolled in schools, some schools attempted to circumvent such requirements by denying access to schools to students based on citizenship status. *Plyler v. Doe* addressed a Texas law that would cease to provide funding to schools to educate students who were undocumented immigrants. The Supreme Court ruled that school districts could not deny entry into schools to any students, regardless of citizenship status. Schools are not permitted to require students and families to provide information about citizenship status.

ESEA Reauthorization. With requirements to provide comprehensible instruction in English to students deemed Limited English Proficient, schools were tasked with how to best serve students. Title VII of the Elementary and Secondary Education Act of 1994 authorized the

Bilingual Education Act, granting a small amount of federal funding to districts with large populations of non-native English speaks. The law favored bilingual education as transitional; that is to say, bilingual education was defined as programming that moved students from being monolingual speakers of one language (such as Spanish) to being speakers of a target language, English, with little to no regard for maintenance of the native language. This subtractive approach to bilingualism privileged the use of English in schools, using students' native language only if needed. Garcia (2014) cites this approach, along with the grudging approach to including Spanish as a foreign language, to guarantee "that monolingual English speakers conserve advantages" (p. 59) in schools across the country.

Evolving views on bilingual education. May (2014) describes two distinct approaches that societies may have toward bilingual citizens' rights to preserve their native language. Tolerance-oriented rights allow citizens to develop language on their own, in private, and without support or discouragement. Promotion-oriented rights utilize a minority language in a public domain, such as in legal or education settings. This may include making some documents available in both languages, or at most, allowing self-governing, such as Native American schools promoting instruction in their language.

Many immigrant groups have sought to preserve their heritage languages over the decades. German/English bilingual schools were common in German communities throughout the country before World War I. In the nineteenth century, language minority groups offered heritage language programming or instruction in Chinese, Japanese, French, Cherokee, Swedish, Danish, Norwegian, Italian, Polish, Dutch and Czech (Lee & Wright, 2014). Unfortunately, not all minority language groups enjoyed even a tolerance-oriented approach. In some areas, Native American students were forced into English-speaking boarding schools and forbidden from using

their native languages. In addition, many regions prohibited the teaching of German after World War I (Lee & Wright, 2014).

No Child Left Behind. A tolerance-oriented approach to bilingualism, with a narrowing focus on English as the only language of value won out in 2001 with the passage of No Child Left Behind (NCLB), which replaced the Elementary and Secondary Education Act. Title III of No Child Left Behind created new requirements for districts to annually assess English learner students for progress and proficiency in the English language, plus content knowledge in math and reading. Schools would be evaluated via Annual Measurable Achievement Objectives, set by each state, to ensure that students made adequate progress in English. NCLB moved the assessment of students and the evaluation of programs to focus entirely around English, without regard for students' native language. In fact, after the enactment of NCLB, many districts reported a funding cut from Spanish courses in order to provide additional funding for math and reading classes (Dillon, 2010). Though NCLB allows students to take tests in their native languages under specific circumstances, a focus on math and reading proficiency has continued to drain resources and focus away from bilingual education (Lee & Wright, 2014).

The decisions to focus on subtractive bilingual education and to cut funding to foreign language programming both contrast with supporting English learner students. Thomas and Collier (2002), found that supporting and developing a student's first language was the best predictor of academic achievement for English learners. Decisions to cut funding to programs supporting students' native languages are, then, wholly counterproductive academically. Powers (2014) challenges schools to recognize language not as a barrier to overcome, but as an excellent resource that English learner students can access.

In December of 2015, Congress passed the Every Student Succeeds Act (ESSA) to replace NCLB. Some noteworthy changes included moving Annual Measurable Achievement Objectives (AMAOs) to Title I instead of Title III, placing the emphasis for language progress and proficiency onto the center stage. The term Limited English Proficient (LEP) has also been eliminated, in favor of the more positive English learner. Two subgroups have been added under English learners, including EL students with disabilities and long-term ELs. Such changed indicate a recognition of the nuanced issues that affect EL students. Unfortunately, critics have been quick to point out a strong emphasis on proficiency in English, with little focus placed on multilingualism, along with a lack of support for teacher education (Mitchell, Robertson & Trez, 2016; TESOL, 2015).

Laws and regulations surrounding the education of English learner students have certainly indicated a pathway toward progress over the years, from segregated Mexican schools to integrated communities. But, although such progress is admirable, there is much work to be done as experts work tirelessly to ensure true equity for English learners. Modern issues facing immigrant students and their families, including overcoming language barriers in the age of accountability (Sheng 2011; Wright & Li, 2008), poverty (Winsler et al., 2012), parents with little to no formal education (Crosnoe & Turley, 2011; Schaller, Rocha & Barshinger, 2007), the stress of the immigration process for recent arrivals (Yoshikawa & Kalil, 2011; Tillman, Guo & Harris, 2006), less access to social programs such as Medicaid, food stamps and pre-k (Karoly & Gonzalez, 2011; Fix & Zimmerman, 2001), and the over-identification of English learners for grade retention (Warren, Hoffman & Andrew, 2014; Xia & Kirby, 2009; Rueda & Windmueller, 2006) leave much to be desired in ensuring that EL students have every chance to receive equitable educational opportunities. Educational leaders and school districts must address as

many of these social and structural issues as possible to provide equitable treatment for all students. Grade retention, with its impact affecting EL students unequally, must be addressed.

In the next section, the history and issues associated with grade retention will be examined to shed light on how this issue has come into significance in modern schools.

A History of Grade Retention

In contrast to the growth and development of programming and placement for English learner students over the decades, grade retention has seen more of a tug-of-war battle among educators supporting or opposing it; consequently, a game of tug-of-war is a fitting illustration of the progress made on the same issue. Many issues of longevity and popularity within the grade retention debate are unresolved in practice.

Post-Civil War to Early 1900s

Toward the end of the Civil War, local schools began to separate students into separate grades with separate curricula and targets for each, immediately raising the issue of what to do with students that did not meet the requirements of their current grade (Holmes & Matthews, 1984). Leonard Ayres, a public school administrator, published his book *Laggards in our Schools* in 1909, addressed the issue:

> In every school there are found some children who are older than they should be for the grades they are in. These children constitute serious problems for the teachers. They are misfits in the classes, require special attention if they are to do satisfactory work and render more difficult the work with other children. They are found in all school systems but are by no means equally common in all systems. (p. 3)

Ayres describes the variability as approximately 7% retained in Medford, Massachusetts, to 75% in Memphis, Tennessee (Ayres, 1909). Ayers found that students were held back due to a lack of skills in English, physical defects, poor teaching, and mobility.

Early Retention Research

A century later, the same questions persist. Many studies have explored rates and causes of retention, and researchers eventually started to question outcomes. Holmes and Matthews (1984) conducted a review of the early literature surrounding grade retention. Empirical studies as early as 1925 tracked retained students, matching subjects over time to attempt to understand the outcomes of grade retention. Of 44 total studies between 1929 and 1981 in the United States and Canada, most studies took place in the 1960s and 1970s, and the studies overall showed a negative effect on students (Holmes & Matthews, 1984). This research was organized into four main themes. Studies focusing on academic achievement explored the math and reading success of students after retention. The average effect size for academic achievement was -0.44, meaning that students had less academic success after retention. Studies also explored students' personal adjustment post retention, with an average effect size of -0.27. In addition, the authors found an average effect size of -0.19 for self-concept, and -0.16 for attitude (Holmes & Matthews, 1984). It was apparent from early empirical research that grade retention was academically and socially counterproductive. Given such a body of literature opposing the practice, one might expect that retention was hotly contested in schools. However, administrators questioned the design and methodologies of the studies as they continued to retain students, sometimes in droves. Though rates of retention had begun to decline in the 1970s and 1980s, the practice had once again begun to pick up. One example listed was Greensville County, Virginia, where 1300 of 3750 students were retained as part of a policy to base grade

promotion on skill mastery. A district in Washington DC reported retaining half of all first, second and third graders due to math and reading proficiency (Holmes & Matthews, 1984).

A Nation at Risk

In the era after *A Nation at Risk*, an estimated 21.3% of adolescents were retained in grade (Resnick et al., 1997). Bowman (2005) describes the transition of social promotion, the practice of promoting students that have not met all requirements for the grade, as common in the 1970s to added pressure for skill mastery in the 1980s. Communities and families demanded that school administrators and teachers hold students to the highest standards. During this time, referrals for special education programs spiked drastically as educators scrambled to maintain high expectations for all students (Bowman, 2005). Such referrals were useful because schools were not held accountable for students in special education programming; for this reason, a referral could help preserve a school's reputation.

A Push for Accountability

During the 1980s and 1990s, a shift from social promotion to a focus on accountability saw increased retention rates, particularly among younger students (Reschely & Christenson, 2013), despite abundant evidence against such the practice and the fact that it is costly to states to educate students for an additional year. Great pressure was being placed on schools to hold students and teachers accountable. In both his 1998 and 1999 State of the Union speeches, President Clinton mentions an end to social promotion in schools:

> We must also demand greater accountability. When we promote a child from grade to grade who hasn't mastered the work, we don't do that child any favors. It is time to end social promotion in America's schools. Last year, last year in Chicago, they made that decision, not to hold our children back but to lift them up. Chicago stopped social

> promotion and started mandatory summer school to help students who are behind to catch up. I propose, I propose to help other communities follow Chicago's lead. Let's say to them, stop promoting children who don't learn and we will give you the tools to make sure they do. (1998)

The next year, President Clinton further outlined a five step plan for accountability for schools, including greater accountability for teachers, accountability for districts' lowest-performing schools, and social promotion:

> Later this year, I will send to Congress a plan that for the first time holds states and school districts accountable for progress and rewards them for results. My Education Accountability Act will require every school district receiving federal help to take the following five steps: First, all schools must end social promotion. Now, no child, no child should graduate from high school with a diploma he or she can't read. We do our children no favors when we allow them to pass from grade to grade without mastering the material. But we can't just hold students back because the system fails them. (1999)

President Clinton's words no doubt sent mixed messages to schools. Though he mentioned mandatory summer school for failing students, specifically referring to Chicago's successful programming, a culture of accountability and reform was already stirring. No Child Left Behind affirmed this in 2001, as did President Bush in his own State of the Union address:

> Critics of testing contend it distracts from learning. They talk about teaching to the test. But let's put that logic to the test. If you test a child on basic math and reading skills, and you're teaching to the test, you're teaching math and reading. And that's the whole idea. (2001)

A new push for accountability from new legislation, public pressure, and the very words of the president pushed the country into even higher rates of retention in the years after the passage of No Child Left Behind (Warren, Hoffman & Andrew, 2014). Schools retained for two primary reasons (Goos, Van Damme, Onghena, Petry & Bilde, 2013). First, repeating a grade allowed a student time to catch up to peers, refresh basic skills and attain others during the second round. Second, having students of weakest academic skills repeat a grade results in, at least, temporarily, homogenous classes. This had the appearance of an effective way to teach students, though the concept was based upon the assumption that students who repeated a grade would continue to make gains after the retained year.

Conclusion

Grade retention has evolved to be a heated issue among scholars and practitioners due to a push for accountability, while progressive practitioners and educators work to design programs to better serve English learners. Two roads, grade retention and the education of bilingual students, converge to indicate a nuanced, complex issue facing many students, families, teachers and administrators in schools today.

Many perspectives on retention of general education students have been enumerated, including academic, social and long-term effects. Such perspectives are worthy of consideration because they paint a picture of the short- and long-term effects of retention on English learner students. Though the present study only seeks to discover the language outcomes of English learner students, the broader picture matters. If an educator were to only consider that academic outcomes are neutral or negative in the long term, it might be reasonable to risk to ensure that a child's formative years are successful; however, knowing that students' socio-emotional and

long-term outcomes are impacted negatively, and that this is clear in the literature, should provide a much clearer picture to educators of the effects of retention decisions.

Chapter 2: Literature Review

A Clustering of Disparities

English learners, and particularly Latino families, who make up a majority of ELs in the United States, experience a "clustering of disparities" (Crosnoe & Lopez Turley, 2011, p. 145) as children, even before arriving at a school. Recently-arrived immigrants from Latin America often describe a hope for their children to have a better life as a reason for immigrating, but upon arrival, the immediate strain of living in the United States puts an incredible amount of stress on their own familial relationships (Bacallao & Smokowski, 2007). These strains include constant fear from immigration officials, financial strain over the process of immigrating, and adjusting to new surroundings (Yoshikawa & Kalil, 2011). Children in immigrant families are less likely than their peers to receive economic assistance such as food stamps and Medicare, even if they qualify, due to a lack of familiarity with the system, language barrier, and/or distrust in the system. Thus, English learner children are likely to enter school with many factors already against them.

Upon arrival to school, the hardships continue. By definition, EL students are not proficient in English (Krashen, 2015). Students are at a disadvantage due to learning content and language simultaneously. This is no small disadvantage; learning English takes five to seven years, or up to ten for students who did not attain literacy in their first language (Varela, 2010; Cummins, 1981). Students are constantly working to understand content presented in a language they do not fully understand. It should be of little surprise, then, that Thomas and Collier (2002) found bilingual education to be the single strongest predictor of academic achievement for EL students, more so than pull-out instruction or other intervention. Bilingual education decreases

the rates of both grade retention and the incidence of dropping out of high school for English learners (Curiel, Rosenthal & Richek, 1986).

A crucial support to EL students for decades, bilingual education took an unfortunate hit with No Child Left Behind, which replaced the Bilingual Legislation Act with legislation for English language acquisition, removing even the word bilingual from federal legislation. Testing for EL students would now be in English only, and schools would be judged on English progress and proficiency, with no set goals or measures for promoting other languages or assisting students in maintaining their home languages. Because students are only tested for English progress and proficiency only, a student's first language became an afterthought or forgotten completely (Menken, 2010).

In addition to lack of first language support, EL students, particularly Latinos, are more likely to feel alienated in school. Sox (2009) describe numerous events recounted by students that led them to drop out of high school, including discrimination by school peers, teachers, and even district administrators. Students reported difficulty or even being prohibited from enrolling, questioning of their immigration status, a lack of interpreters for families, few to no English language development programs, and all-around poor treatment. Students, particularly English learners, may leave due to feeling marginalized (Sox, 2009).

The problem goes beyond discrimination, however. NCLB requirements for high test scores create enormous pressure for administrators and teachers to produce high achievement numbers. EL students as a category produce lower test scores due to the very complex and difficult language of a test in English (Wright & Li, 2008). Linda Darling-Hammond (2006) describes the effects of such as testing culture:

> NCLB's requirement for disaggregating data and tracking progress for each subgroup of students increases the incentives for eliminating those at the bottom of each subgroup, especially where schools have little capacity to improve the quality of services such students receive. The consequences for students who are caught in this no-win situation can be tragic, as most cannot go on to further education or even military service if they fail these tests, drop out, or are pushed out to help their schools' scores look better. The consequences for society are also tragic, as such policies lead to more students leaving school earlier — some with only a seventh-or eighth-grade education — without the skills needed to participate in the economy (Darling-Hammond, 2007, p.251).

Hostility in schools and pressure to leave are perhaps why Hispanic students have much higher dropout rates than their peers (NCES, 2014). Moreover, EL students are more likely to be retained than non-EL peers of the same achievement levels (Tingle, Schoeneberger, & Algozzine, 2012; Xia & Kirby, 2009; Willson & Hughes, 2006). It is well-established in scholarly literature that elementary grade retention causes students to drop out of high school (Hughes, West, Kim & Bauer, 2018; Hughes, Cao, West & Smith, 2017; Braun, Gable, Billups, Vieira & Blasczak, 2016; Allensworth, 2005; Jimerson, 1999; Roderick, 1994). An important study by Stearns (2007) indicated that, while all students who were retained experience a higher likelihood of dropping out, retained Latino students were 24 times more likely to drop out of school than their promoted counterparts. Modern studies continue to support this assertion: a new longitudinal study has shown that one of the strongest outcomes after retention is dropping out of high school, and this is particularly true for Hispanic girls (Hughes, West, Kim, H. & Bauer, 2018).

Despite the very real threat of causing students to drop out of high school, retention continues to happen en masse across the United States each year. Characteristics of some, but not all ELs, put them at additional risk for being retained. For example, research indicates that EL students with low literacy scores are more likely to be retained (Goldstein, Eastwood, & Behuniak, 2014; Winsler et el., 2012). Students are recategorized as former-EL once they secure high enough scores; for that reason, by definition, EL students have low, non-proficient literacy (reading and writing) scores. In addition, students whose mothers have lower levels of education are also more likely to be retained (Byrd & Weitzman, 1994). Schaller, Rocha and Barshinger (2007) report that 91% of the population of Mexico in the 1970 did not complete grades 1-6; if this holds true for later generations, it is likely that many EL students have parents with lower levels of educational attainment. Furthermore, immigrant children are less likely to attend pre-kindergarten due to affordability and availability of programs (Karoly & Gonzalez, 2011), despite the many benefits, including early English skills, social resources, and even a decreased likelihood for later grade retention (Lazarus & Ortega, 2007; Ramey & Ramey, 2004; Gilliam & Zigler, 2000). Finally, families of EL students may be culturally less likely to fight against school recommendations of retention. String (1960) identifies parental involvement as the difference between retention and social promotion; that is to say, parents that hotly contest retention are likely to have their students socially promoted, despite academic achievement at the time of the decision. Willson and Hughes (2006) found this to be true: a low sense of responsibility in the decision was a predictor of retention. In her book *The Culture Map* (2014), Erin Meyer describes various aspects of world cultures and how they relate to decision-making. When it comes to leadership, the culture of the United States tends to value egalitarian decision-making over hierarchical structures; however, countries such as Mexico, China, India, Japan,

Korea and Russia value hierarchical decision making. Families from countries with these deep cultural beliefs are less likely to dispute a decision recommended by the school.

The clustering of disparities described by Crosnoe and Lopez-Turley refers to the many hardships that immigrant student face upon arrival to the United States. The research findings synthesized in this section explain the many disparities that immigrants or the children of immigrants face when they arrive in school. Unfortunately, there a lack of literature to shed light on the exact outcomes of grade retention for English learners. The next section will detail scholarly work surrounding achievement outcomes for all students, but little exists to inform educators of the academic outcomes of retention for ELs. Retention decisions have serious and long-ranging consequences for students, particularly for ELs; therefore, it is the onus of the educator to be informed, and of the scholar, to inform practitioners.

Grade Retention and Academic Achievement

The problem of remediating struggling students has been the subject of many scholarly investigations, in an attempt to determine the most effective way to meet students' needs. For well over a century, educators and administrators have utilized the repetition of a grade for some students who fail to meet the requirements of the grade they are in. Academic achievement, or lack thereof, is the primary reason that students are retained in grade (Bowman, 2005; Tomchin & Impara, 1992; Rafoth 1991; Abidin, Golladay, & Howerton, 1972; Ayres, 1909), with one study reporting that academic performance and ability level comprised 78% of reasons given for retention by teachers, and 72% of reasons given by school administrators (Range, Pijanowski, Holt & Young, 2012). It is entirely reasonable, then, that academics have pored over the question of the effectiveness of grade retention as an academic intervention for so many decades

(Vandecandelaere, Vansteelandt, De Fraine & Van Damme, 2016; Reynolds, 1992; Keyes, 1911).

Research design for grade retention studies. Classic study design in the scholarly literature includes a research question about outcomes for students who were retained as compared to other students. Outcomes investigated are often reading and/or mathematic test scores in studies conducted in the United States (Mariano & Martorell, 2013; Moser, West, & Hughes, 2012; Hughes, Chen, Thoemmes & Kwok, 2010; Burkam, LoGerfo, Ready & Lee, 2007; Ferguson, 1991; Dobbs & Neville, 1967), while studies conducted in other countries may include achievement in foreign language and science (Ehmke, Drechsel, & Carstensen, 2010; Klapproth et al., 2016). Retention is utilized as a treatment variable and achievement before and after the repeated grade is measured. Some studies utilize state assessments or scores that are already available (Dong, 2010; Cooley, Navarro & Takahashi, 2011) while other utilize an independent assessment such as the Woodcock-Johnson (Chen, Hughes, & Kwok, 2014; Wu, West & Hughes, 2008; Jimerson, Carlson, Rotert, Egeland & Sroufe, 1997).

Selection is of the utmost importance in scholarly studies, and is often the subject of debate among scholars investigating grade retention. Random assignment of students into retention groups would be unethical and surely unpopular with parents; for this reason, researchers must find ways to compare groups that are inherently unlike. Early selection for groups was well-intentioned but limited. For example, one study compared 85 retained students with 43 students who were promoted but scored in the 25th percentile on a readiness assessment (Abidin, Golladay & Howerton, 1972). While this did match students before comparing groups, the assumption that students were retained due to readiness was integral to the study design. Other designs have taken into consideration promotion standards, comparing students who were

retained due to promotion standards with students who barely made the cut (May & Welch, 1984; Ferguson, 1991; Wang & Johnstone, 1997), with the rationale that group achievement was similar. However, this cannot take into account differences in achievement, such as students who excel in math but struggle in reading, nor can it account for background variables found to be significant in similar educational studies. A pattern began to emerge in scholarly literature that researchers would match students on a number of variables in order to better compare groups. Scholars went beyond single-measure assessments and began to take into account a more holistic view. Earlier studies matched students on a number of variables, including race, sex, age, and achievement and then did simple comparisons of outcomes (Dobbs & Neville, 1967), while later investigations also employed regression to statistically adjust for performance before retention (Reynolds, 1992; Jimerson, Carlson, Rotert, Egeland & Sroufe, 1997; Jimerson, 1999; McCoy & Reynolds, 1999).

In 2005, Roderick and Nagaoka released a study documenting outcomes after retention in a design that stood out from previous methodologies. The authors praised previous studies for good approaches to matching students to create groups, but simultaneously called for an improved approach based on several arguments. First, matches based on a few variables could not account for regression toward the mean; that is to say, it would be hard to determine whether a retained student who had a particularly bad year and would likely bounce back the next year, could be compared to a student who was promoted and perhaps had a particularly good year, from which he or she might bounce back in the other direction. Second, selection based on a promotion threshold, while effective for removing selection bias of educators, cannot account for students' lower or higher capacity to respond; for example, a higher-achieving student might have become nervous and performed poorly on a single-measure assessment, whereas a lower-

achieving student might have gotten lucky on a few answers. Thus, the authors called for more refined methodology in selecting matched groups. The matching they offered provided additional variables and was arguably more nuanced than prior inquiries; however, the challenge and debate to ensure better-matched groups continued.

A 2006 study introduced a more advanced concept for creating groups in retention studies. Silberglitt, Appleton, Burns and Jimerson toyed with propensity score matching as a technique for creating matched groups. Though it was not used in the final analysis, it was employed as a tool for ensuring comparison groups were alike.

Rosenbaum and Rubin were some of the earliest authors on propensity score matching (PSM):

> In randomized experiments, the results in the two treatment groups may often be directly compared because their units are likely to be similar, whereas in nonrandomized experiments, such direct comparisons may be misleading because the units exposed to one treatment generally differ systematically from the units exposed to the other treatment. Balancing scores, defined here, can be used to group treated and control units so that direct comparisons are more meaningful (1983, p. 42).

In its genesis, PSM was used primarily in medical research to combat the issue that Holland (1986) described as the "fundamental problem of causal inference" (p. 947) in observational studies; that is, the trouble of validity in comparing two nonrandom groups. Rosenbaum and Rubin give the example of analyzing treatment of heart disease with coronary bypass and drugs. Treatment naturally precludes random assignment. Because it would be unfair and nonsensical to compare outcomes of those having coronary bypass with those who did not

need it, the authors propose statistically matching individuals undergoing each treatment in order to create like groups.

Though PSM was used first in medical research, the approach was eventually picked up by educational researchers investigating grade retention and quickly became the standard for retention research. After an early mention in retention research in 2006, Hong and Yu (2007) employed the technique to determine causality of grade retention as related to academic outcomes, arguing that promoted and retained groups are so inherently different because "most promoted children have little or no risk of ever being retained. When the two groups are barely comparable, statistical adjustment for a limited number of background variables cannot be relied upon to remove bias" (p. 409).

In their breakthrough study, Hong and Yu investigated kindergarten and first grade retention, as compared with continuous promotion, finding that both had immediate negative effects on students' performance in math and reading; however, effects faded by fifth grade. Retained students were no better or worse off than continuously promoted matched peers, except that the retained students were one year older, effectively having lost a year (Hong & Yu, 2007). Studies have utilized very similar approaches to research questions and methodology in recent years, comparing retained, matched promoted, and a random group of students. From 2008 to 2012, several studies investigated post-retention achievement (reading or math) outcomes. During these years, fewer studies utilized propensity score matching (Dombek & Connor, 2012; Peterson & Hughes, 2011; Ehmke, Drechsel, Carstensen, 2010; Dong 2010; Chen, Liu, Zhang, Shi & Rozelle, 2010; Wu, West & Hughes, 2008), while other studies during this era used different statistical procedures to analyze data over time (Tingle, Schoenberger, Algozzine, 2012; Cooley, Navarro & Takahashi, 2011; Martin, 2011; Martin, 2009). Since 2012, a higher

percentage of studies have utilized PSM (Fruehwirth, Navarro & Takahashi, 2016; Klapproth et al., 2016; Vandecandelaere, Vansteelandt, De Fraine & Van Damme, 2015; Lamote, Pinxten, Van Den Noortgate & Van Damme, 2014; Mariano & Martorell, 2013; Goos, Van Damme, Onghena, Petry & de Bilde, 2013; Raffaelle Mendez, Kim, Ferron & Woods, 2015). It is perhaps indicative of the growing popularity of the technique that very few studies regarding academic achievement and grade retention have been published since 2014 without the use of PSM.

Findings of retention studies. Though sophisticated statistical matching has been in vogue for retention studies for only a decade, researchers have matched students to some degree for many years. The earliest retention studies indicate a dire warning against retaining students (Keyes, 1911), and the same warnings have resonated through the decades (Cooley, Navarro & Takahashi, 2011; Reynolds, 1992; May & Welch, 1984; Abidin, Golladay & Howerton, 1972) yet grade retention continues to be prevalent across the country (Warren & Saliba, 2012; Shepard & Smith, 1990), primarily for the purpose of boosting achievement (Range, Pijanowski, Holt & Young, 2012). One insight is that teachers do not have, or choose not to access, the scholarly literature, with one study finding that only 9% of teachers attributed their knowledge about grade retention to scholarly work (Witmer, Hoffman & Nottis, 2004); however, confusion about retention outcomes may also be attributed to the fact that research does not point one single direction on the issue. Findings on retention are complex and nuanced. This section will explore such mixed results of grade retention studies exploring academic outcomes.

Support for Retention. Studies have indicated concerns surrounding grade retention and its effectiveness for a number of years; however, a few point in a seemingly opposite direction. Three studies indicated that retention was actually beneficial to students. A study of a policy adopted in Florida in 2002, one of the first of its kind, with many states to follow, investigated

the effectiveness of retention on reading scores (Greene & Winters, 2007). The state's policy required third graders to achieve a Level Two (of five) on a standardized reading test in order to move on to fourth grade. Researchers tested promoted students that would have been retained the year before the policy went into effect, as compared with students who were retained the first year of the policy, but without matching background variables, finding that students retained in third grade were more likely to pass fifth grade reading tests than promoted students of similar achievement pre-retention (Greene & Winters, 2007). Two red flags add the need for caution when drawing conclusions about this study. First, according to the authors, the demographics we "statistically significant, but arguably insubstantial" (p. 330). Statistical matching was not utilized, leaving unmatched groups for comparison. Of greater concern is the fact that students retained under Florida's policy had interventions and additional treatments in the retained year, adding serious limitations to any conclusions that the author or reader may draw. Still, the study emerged as evidence that Florida's policy was working to help students pass fifth grade tests.

Two additional studies have emerged to indicate benefits to retaining students. A study of grade retention in Texas held many similarities to the aforementioned Florida study, investigating the probability of passing standardized exams after retention (Hughes, Chen, Thoemmes & Kwok, 2010). The authors found a positive association between retention in first grade and passing state reading and math exams in third grade, despite limitations that over 25% of third grade assessment data was missing for retained students. Also concerning, though not listed as a limitation, is the assertion that retained students likely received additional interventions. An extra year of instruction certainly should count as intervention, as students were compared in third grade. Regardless, the study indicated that retained students were around 1.8 times more likely to pass reading and math assessments in third grade, as compared with

matched-promoted peers. A second study illustrating an academic benefit to retention tracked students retained in kindergarten. An analysis of test scores indicated that students repeating kindergarten fared better on first grade reading and math assessments, and by third grade benefits were still evident in math, but not reading (Dong, 2010).

The results from the three studies appear to demonstrate consilience over the retention issue; taken in an isolated context, a practitioner might conclude that the evidence is quite clear. However, one small aperture presents itself in the third study, which raises important questions. Students retained in kindergarten performed better two years (first grade) and four years (third grade) after retention in math, but progress in reading, evident in first grade, had disappeared by third grade, begging the question of what happened in third grade. The other studies supporting retention (Greene & Winters, 2007; Hughes, Chen, Thoemmes & Kwok, 2010) only followed students for two years. What conclusions might be drawn if studies were extended beyond this two-year scope? In fact, this is the entirety of the difference between studies supporting retention versus those concluding that it is not beneficial or even harmful. The next section will explore the plethora of longitudinal studies exploring academic outcomes of retention.

Longitudinal studies before PSM. As early as the 1950s, studies began to examine the outcomes of retention for matched groups of students. An early study compared retained and promoted students in the same school, finding that students made more progress when they were promoted, as compared with retained students (Coffield and Blommers, 1956). In the next decade, a similar study sought to answer the same question, but first matched retained and promoted students on a number of variables: race, sex, socioeconomic status, age, reading achievement and mental age (maturity), tracking matched students over time (Dobbs & Neville, 1967). Findings were similar, indicating that promoted students achieved more over time, but

perhaps a greater revelation of this study was the idea that matching students on basic variables was a key part of study design. Studies conducted after this consistently matched students on background variables.

A Nation at Risk (1983) sparked a heated debate on the rigor of public schools, and soon after more literature arose, enlightening practitioners about the practice of grade retention and its academic outcomes. The 1990s and early 2000s produced a number of retention studies, each utilizing some approach to match background variables statistically. Some studies found a one-year boost in student achievement for repeat year, with the effects disappearing almost immediately after (Jimerson, Carlson, Rotert, Egeland & Sroufe, 1997; Rust & Wallace, 1993). A longer-term study found that students retained in an early grade showed significantly lower reading at age 14, as compared with a matched-promoted peer (McCoy & Reynolds, 1999), while another study found that retained students were more likely to be referred to Special Education (Roderick & Nagaoka, 2005).

Longitudinal studies utilizing PSM. The majority of studies utilizing advanced statistical techniques have indicated negative outcomes after retention. Two exceptions exist: in some cases, researchers employing PSM have found short-term positive outcomes post-retention (Mariano & Martorell, 2013; Hughes, Chen, Thoemmes & Kwok, 2010). Yet, when closely examined, these studies are actually consistent with the findings of other modern literature. Another study in the same era (Dong, 2010) found that students indicated a boost in both math and reading after the retained year (kindergarten). Students maintained gains in math, though gains in reading leveled off by third grade, the end of the study. A study two years later might shed further light on the same question: students retained in first grade, despite presenting an initial boost, arrived to fifth grade with somewhat lower reading achievement and significantly

lower math skills when compared with matched peers (Moser, West & Hughes, 2012). An investigation of secondary students in Luxembourg mirrored these outcomes; students experienced a short-term boost which dissipated entirely over time (Klapproth et al., 2016). Thus, the initial boost which is evident in a number of studies does not paint the entire picture; upon examination over many years, any initial benefit appears to fade into detriment.

Other studies utilizing PSM failed to find even a short-term benefit; instead, students seemed to encounter only negative academic outcomes post-retention. Two studies investigated retention in kindergarten and first grade, finding that negative effects were immediately apparent but dissipated through the elementary years (Vandecandelaere, Vansteelandt, De Fraine & Van Damme, 2015; Hong & Yu, 2007). Chen, Liu, Zhang, Shi and Rozelle (2010) also found only negative outcomes for retained students, particularly for students retained in second grade. Similarly, Wu, West and Hughes (2008) reported a decrease in math achievement in the years after first graders were retained, though reading achievement was not significantly different. Im, Hughes and Kwok found that reading and math achievement was no different for retained students than for matched-promoted peers (2013), and Silberglitt, Jimerson, Burns and Appleton found that retention was harmful for both early elementary (kindergarten through third grade) and late elementary (fourth through sixth grade), albeit more so for later grades.

An abundance of evidence indicates that grade retention may achieve the precise opposite of its purpose; instead of preparing students for future success, the literature shows that the effect is negative or, at best, neutral, in addition to setting students behind a year.

Other academic factors. Beyond test scores, scholarly inquiries have shown that grade retention affects students' performance in school in a number of ways. Teachers reported that retained students struggled more to pay attention, as compared with matched-promoted peers,

perhaps because they are experiencing the same content for a second time (Raffaele Mendez, Kim, Ferron & Woods, 2015). Along the same lines, retained students are (incorrectly) perceived by teachers to be less academically competent by their teachers (Dombek & Connor, 2012; Reynolds 1992). Students describe their repeat year experiences with frustration, as being the same teacher, the same classroom and the same curriculum (Stone & Engle, 2007).

Differences in student behavior are also apparent after grade retention. Researcher Shane Jimerson has conducted extensive research on this topic, finding that students were more likely to demonstrate chronic absenteeism and maladaptive behaviors (Jimerson, Carlson, Rotert, Egeland & Sroufe, 1997), academic adjustment (Jimerson, 1999) and more aggressive behaviors (Jimerson & Ferguson, 2007; Rodney, Crafter, Rodney & Mupier, 1999) over time as compared with matched peers. Other studies have found that students have lower self-esteem (Martin, 2011; Carlton & Winsler, 1999), lower confidence in academic tasks (Huddleston & Lowe, 2014), higher levels of disengagement (Martin, 2009) and, in the case of later retentions, more difficulty making friends (Demanet & Van Houtte, 2016).

Retention and English learners: A Clustering of Disparities

The "clustering of disparities" described by Crosnoe and Lopez Turley (2011, p. 145) was meant to describe the ascendancy of difficulties faced by new immigrants upon their arrival to the United States. However, the term is as applicable as ever to English learners' school experiences. A number of factors make school the issue of grade retention a near emergency for this population.

EL students as a population come from diverse backgrounds and school experiences, but they share one commonality: by being EL, they are defined by a lack of proficiency in English and are therefore particularly vulnerable to certain hardships of school. First, English learners are

statistically more likely to be chosen for retention. Numerous studies have shown English proficiency to be a predictor of retention (Winsler et al., 2012; Tingle, Schoeneberger & Algozzine, 2012). This may be attributed to a lack of proficiency in English skills, such as reading (Willson & Hughes, 2006) or a lack of proficiency in academic language (Wright & Li, 2008). Demographic characteristics often correlated with EL students also increase the likelihood of retention; for example, Latino students have been found to be targeted for retention more often than other subgroups (Warren, Hoffman & Andrew, 2014; Xia & Kirby, 2009).

Status as an immigrant may also increase the likelihood that a student is retained. First generation immigrants may be more likely to be retained (Tillman, Guo & Harris, 2006). Factors such as the education system in the home country may also influence a family for generations, a phenomenon Crosnoe and Lopez-Turley (2011) statistically tracked to certain regions where instability abounds. For example, half of Mexican immigrant children have no parent with a high school degree, while the majority of East Asian immigrants have at least one college-educated parent (Crosnoe & Lopez-Turley, 2011). According to Schaller, Rocha and Barshinger (2007), 91% of the population of Mexico in the 1970s did not complete sixth grade. Byrd and Weitzman (1994) found high maternal education to be a negative predictor of grade retention, which is to say that more educated parents tended not to have their students retained. A modern-day echo of that study found that more educated parents were more often granted exemptions to retention policies, but students were more likely to be retained if their mother was born in another country (LiCalsi, Ozek & Figlio, 2016).

Immigrant children are less likely to attend pre-kindergarten (Yoshikawa & Kalil, 2011; Karoly & Gonzalez, 2011), but pre-kindergarten attendance has been shown to decrease the risk of retention (Lazarus & Ortega, 2007; Ramey & Ramey, 2004; Gilliam & Zigler, 2000).

Similarly, being a recent immigrant may be correlated with poverty (Raphael & Smolensky, 2009), which is a consistent predictor of being retained (Byrd & Weitzman, 1994; Akmal & Larsen, 2004; Malone, West, Denton, Park & NCES, 2006; McGill-Franzen, Zmach, Solic & Zieg, 2006; Dong, 2010; Winsler et al., 2012).

Another major concern is the likelihood that students will leave school early. Grade retention has long been viewed as a causal factor for students dropping out of school (Mann, 1986; Roderick, 1994; Jimerson, 1999; Jimerson, 2002; Jacob & Lefgren, 2009; De Witte et al., 2013). Bilingual education, an effective intervention for preventing drop outs (Curiel, Rosenthal & Richek, 1986) is less and less common since No Child Left Behind (Garcia, 2014). Perhaps, then, it is no wonder that English learners are more likely to drop out of high school than other populations (Sheng, 2011). In addition, in the case of Hispanic English learners, there is a significantly higher dropout rate for Hispanic males than their peers (NCES, 2014; Stearns, 2007).

If English learners are more likely to be retained, and if retention may result in dropping out of high school, and if the target of the retention may already be more statistically likely to drop out in the first place, then the decision to retain an English learner must be undertaken with great caution. So what scholarly literature informs the decision to retain? Only one study has undertaken the question, and the consideration of English learners was only a component of a larger study. Wu, West and Hughes (2008) investigated reading and math outcomes of students after retention. All students were negatively affected by retention, but English learners in particular experienced a greater slowing in their academic growth. No other studies exist to shed light on the academic outcomes of grade retention for English learners.

It is imperative to better understand the outcomes for English learners after retention in order to inform educators as they make decisions to retain or promote English learners. More information will be crucial to ensure that this diverse and sometimes vulnerable population has the best possible chance for a great education.

Framework for This Study

Given the available literature, the present study will seek to understand outcomes after retention for English learners. Students will be matched on reading and writing proficiency, as these skills are significant for student success in schools (Cook, Boals & Lundberg, 2011) and have been shown to increase chances of retention (Goldstein, Eastwood & Behuniak, 2014; Winsler et al., 2012), as well as other factors that have been proven significant in previous studies. For example, two demographic characteristics may increase a student's likelihood for being retained: being male or Hispanic (LiCalsi, Ozek & Figlio, 2016; Warren, Hoffman & Andrew, 2014; Tingle, Schoeneberger & Algozzine, 2012; Greene & Winters, 2009; Martin, 2009; Burkam, LoGerfo, Ready, & Lee, 2007; Malone et al., 2006; Tillman, Guo & Harris, 2006, Jimerson et al., 2006).

Though the literature on EL students and grade retention is limited, previous work has indicated that retention is likely detrimental to EL students' academic progress (Wu, West & Hughes, 2008). However, this was specific to academic progress (reading and math) and did not necessarily test academic language, though that it a part of reading comprehension (Lesaux, 2012). Still, overwhelming evidence of negative (or at best, neutral) academic outcomes after retention for English-only or non-EL students, or for a student population including ELs indicates that retention is likely not helpful to any subgroup (Fruehwirth, Navarro & Takahashi, 2016; Vandecandelaere, Vansteelandt, De Fraine & Van Damme, 2015; Goos, Van Damme,

Onghena, Petry & de Bilde, 2013; Dombek & Connor, 2012; Cooley, Navarro & Takahashi, 2011; Martin, 2011; Chen, Liu, Zhang, Shi & Rozelle, 2010; Allen, Chen, Willson & Hughes, 2009; Wu, West & Hughes, 2008; Roderick & Nagaoka, 2005; McCoy & Reynolds, 1999; Reynolds, 1992).

In fact, English learners may be more vulnerable to negative outcomes after retention due to the socio-emotional impacts of repeating a grade, including lower confidence in reading (Huddleston & Lowe, 2014; Martin, 2011; Carlton & Winsler, 1999) and more difficulty with social acceptance (Lavrijsen & Nicaise, 2017; Demanet & Van Houtte, 2016; Wu, West & Hughes, 2010) because language learning happens best when students are not experiencing fear. Krashen (1987) calls this phenomenon the affective filter, and when it is activated in a moment of fear, frustration or stress, language learning is inhibited. Therefore, it is entirely possible that English learners experience an interruption in both academic and language development after retention, as has been exhibited in scholarly literature (Mohammadpur & Ghafournia, 2015; Lin, 2008).

This study will contribute to current literature by assessing the impact of grade retention on the language development of English learners. It is imperative to answer the question of whether grade retention may be harmful to an already-vulnerable population that is so susceptible to this treatment.

Research Question

Does grade retention influence language development in English learners, as measured by WIDA Listening, Reading and Writing scale scores, seven years after grade retention?

Null hypothesis: Retention does not influence WIDA 2009 to WIDA 2015 Listening, Reading and Writing scale scores.

Alternative hypothesis: Retention negatively influences WIDA 2009 to WIDA 2015 Listening, Reading and Writing scale scores.

Chapter 3: Methods

Introduction

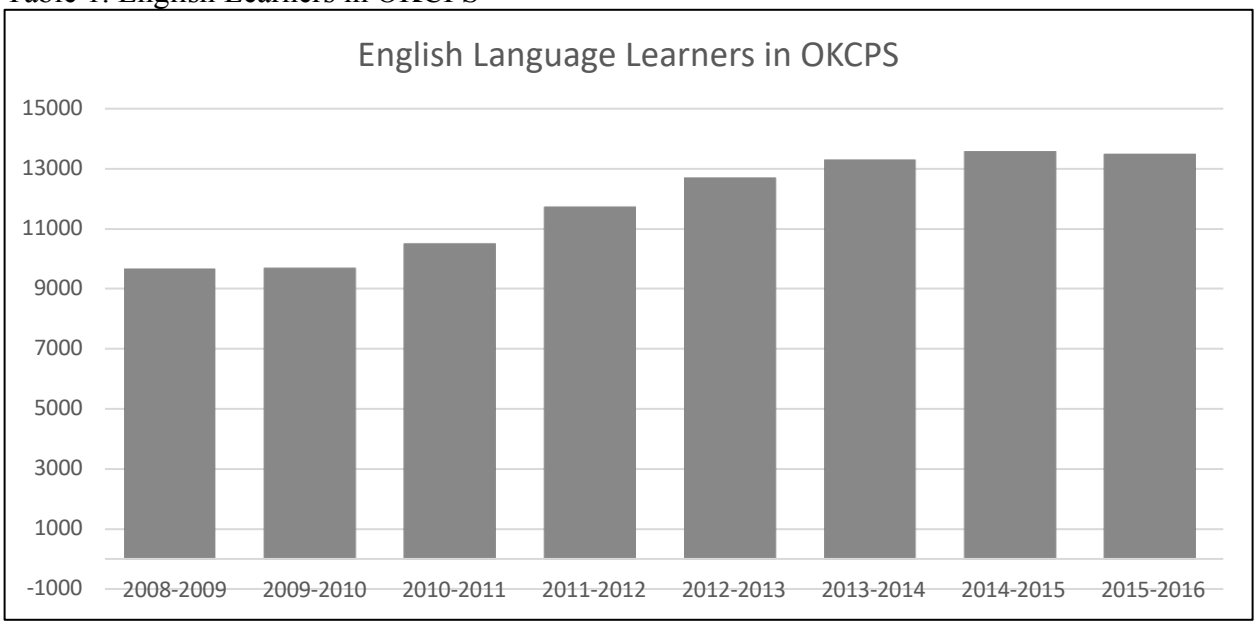
The present study will contribute to a growing body of literature by exploring the language development outcomes of retained students in the years after retention. This chapter will introduce the setting in which the problem is situated, the instrumentation used in the study, variables of significance, study procedure and data analysis.

Setting

Oklahoma City Public Schools serves the highest English learner population in the state of Oklahoma with over 13,000 EL students annually between 2013 and 2016 (OKCPS, 2017). This number has increased each year for the almost-decade that the district reports, as is evident in Table 1 below (OKCPS, 2017; OKCPS, 2014). Tulsa, the next highest student population in the state, boasts around 7,400 EL students annually (TPS, 2017).

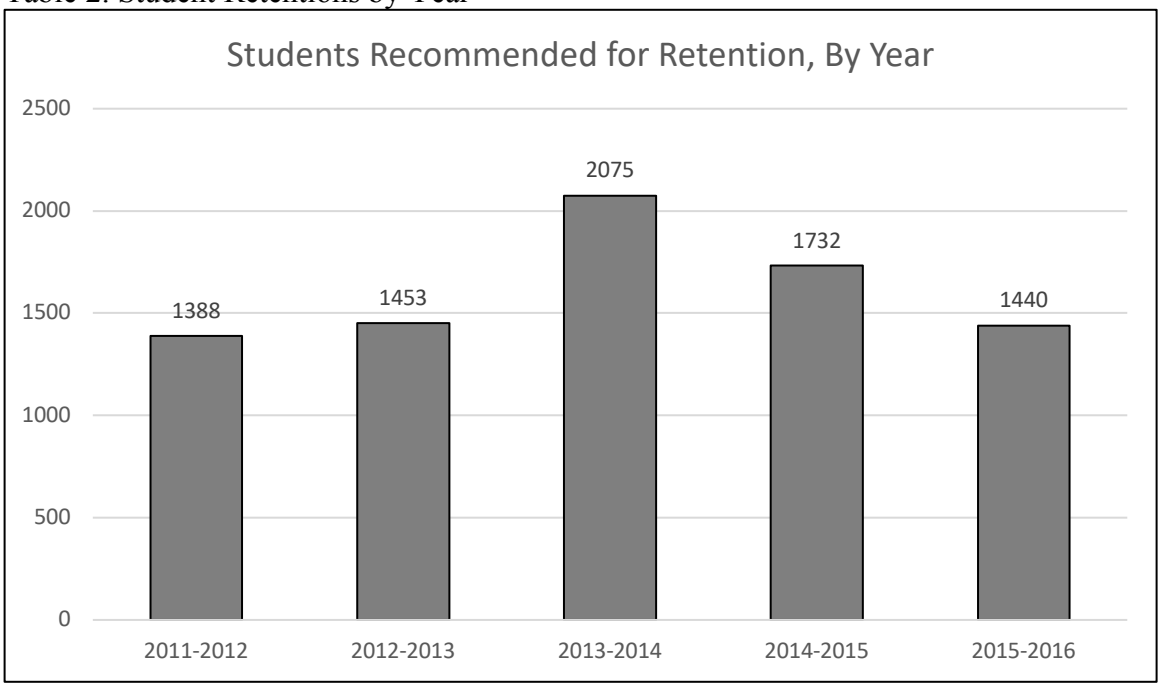
Oklahoma City Public Schools offers varying services to English learners, all of which center around integrating EL students into mainstream classes, to be immersed in the language, while supported in language development. Services include modified instruction provided by mainstream teachers, English language development classes with a specialist, dual language offerings in some sites, support from bilingual paraprofessionals, and some courses taught in students' first language. An orientation toward mainstreaming is widely considered to be a best practice for English development and is strongly supported by research, meaning that most students are served in that context. The mainstream classroom is also where the decision to retain or promote often happens.

Table 1. English Learners in OKCPS



The present study will investigate the language development outcomes of EL students in OKCPS who were retained in the spring of 2009 or 2010, as compared to students who were continually promoted throughout the seven years of the study. Teacher-initiated retention is a strong trend in the district, seemingly on the rise. Recent legislation from the state, enacted in the 2013-14 school year, has also seemed to increase the number of students annually recommended for retention. As of 2018, Oklahoma is second in the nation for grade retention (Palmer, 2018).

Table 2. Student Retentions by Year



A chi square goodness-of-fit test was conducted for grade retention data for five years in the district. Numbers of retained EL and English-only students were compared with proportions of EL and English-only students in the entire district population. EL students are approximately 30% of the entire district population, but represent a much higher percentage of district retentions. The chi square test indicated a significant overrepresentation of EL students being targeted for grade retention, as is evident in Table 3.

Table 3
Chi square goodness-of-fit for OKCPS retention data

	2009-10	**2010-11**	**2011-12**	**2012-13**	**2013-14**
p	<.001	<.001	<.001	<.001	<.001
X^2	136.2571276	191.139593	265.6480231	290.958632	391.1423461
Crit	20.278	20.278	20.278	20.278	20.278
Odds Ratio	2.40	2.48	2.81	3.01	2.71

An odds ratio calculation indicates that for five years, English learner students were 2.4 to 3.01 times more likely to be targeted for grade retention, as compared with English-only peers.

Such findings are consistent with academic literature indicating that English learner students are more likely to be retained (Willson & Hughes, 2006, Xia & Kirby, 2009, Winsler et al., 2012, Warren, Hoffman, & Andrew, 2014).

Identification of English learner students for grade retention likely poses a hazard for a subgroup of students who already struggle in school (Wu, West & Hughes, 2008). Why, then, are EL students still retained so often? First, there is some controversy in grade retention literature: specific studies have shown positive academic outcomes for retained students one or two years after grade retention (Dong, 2010; Mariano & Martorell, 2013). Though a plethora of literature indicates that studies with longitudinal design demonstrate a complete dissipation of the temporary academic boost of retention, controversy about whether or not retention is harmful tends to linger. This is particularly true for Oklahoma City Public Schools, where some individuals in leadership have held pro-retention beliefs. Other individuals doubt the application of studies to the population of Oklahoma City Public Schools as an urban district. Finally, while there is literature regarding EL students and academic (reading/math) outcomes, there is no conclusive academic literature indicating positive or negative language outcomes of retained EL students over time. Educators retaining EL students do not have the information they need for making such crucial decisions. For example, a common perception is that English learner students need more time to learn language before they are pushed into the next grade. More evidence is needed to inform this practice.

Instrumentation

The state of Oklahoma uses the World-Class Instructional Design and Assessment (WIDA) Assessing Comprehension and Communication in English State-to-State for English learners (ACCESS) for ELLs to assessment track English learner student progress and

proficiency according to requirements laid out by Title III of the No Child Left Behind and, subsequently, the Every Student Succeeds Act. Oklahoma City Public Schools' student information system tracks both Proficiency Level and Scale Score results in all four tested domains (Listening, Speaking, Reading and Writing), plus Literacy and Overall Composite scores. The Proficiency Level (PL) score indicates a point on a scale of language proficiency between 1.0 and 6.0, with lowest scores indicating beginning stages of language development. In 2009, a composite PL score of 5.0 indicated proficiency in English according to Oklahoma's AMAO 2. In addition to Proficiency Level scores, a three-digit Scale Score is provided.

Whereas proficiency level scores are relative to a scale of overall language proficiency, scale scores assist educators in comparing student scores from one year to the next. Oklahoma's AMAO 1 indicated a requirement for adequate growth for EL students as 21 scale score points per year; however, such a measure was controversial. Cook, Wilmes, Boals and Santos (2008) indicate that a calculation of student growth should not be so simple; instead, the authors indicate, scale score growth should follow the mantra *lower is faster, higher is slower*, meaning that students of lower starting proficiency can be expected to grower at rates much more accelerated than students of higher starting proficiency. In other words, expecting 21 Scale Score points of a newcomer student that is learning English for the first time might be quite reasonable, but it is not the same as expecting 21 scale score points from an advanced or long-term EL student, for whom 21 scale score points would be a great leap. Because the US Department of Education does not allow a tiered system for the calculation of AMAO 1, a flat expectation of 21 scale score points was utilized; however, it is important for educators to know and understand the nuanced nature of scale scores when calculating growth for students. In a

study design, it is important for researchers to include both Proficiency Level and Scale Score data in order to best group and track students over time.

According to WIDA's technical report (Yanosky et al., 2014), the Access assessment is checked for inter-rater reliability (writing assessment), items are checked for construct validity, normality, skew, and bias. Additional information can be found in WIDA's Access for ELs Technical Reports.

Procedure

Data were collected from the district Student Information System and archives. Only students who were served (current) English learners with seven years of data were kept in the sample (n=942). Students were excluded if they were English only (not served) or if they were served ELs that tested out during the course of the study, as scores were required for each year in the study. Two entries for student grade were included for each year of the study: the grade from the Student Information System, and the test grade when students took the WIDA exam. In some cases, one or the other was missing, so a new variable was calculated to use the mean of the two original data. This was done for each year in the study. Kindergarten was coded as grade zero initially, but later coded as 1 (n=291) as a holdout variable for the statistical model.

Once student grade data were complete, retention status was calculated by subtracting one grade from the other; for example, the grade a student was in for 2010, minus their grade in 2009 (g10-gr09). If a student was in first grade both years, the result would be zero. This variable was then reverse coded so that retained students were coded as 1 for each year. Students who were retained in spring 2009 or spring 2010 were coded as such; many of these students would make up the treatment group. Students retained in spring 2011 or after were coded as such and

eliminated from the sample, as they do not belong in the treatment group (retained in 2009 or 2010) or the control (never retained).

Table 4
Number of Students Retained by School Year

2008-09	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15
84	59	48	43	15	11	1

September first age was calculated by formatting the date of birth variable, creating a new variable to reflect September 1, 2008, and then creating a calculated variable to show the difference between the two (*datediff*), formatted as years. A second calculation determined student's September 1 age relative to his or her grade. The present study defines age relative to grade as a calculation derived from the student's age on September 1, 2008, minus the grade, minus five. For example, a student with the September 1 age of 7 in first grade would be +1 for his/her grade ([7-1]-5)=1, indicating that the student is one year over the typical age for grade. No students in the sample were younger than expected for their grade. Calculations were checked by a review of descriptive statistics and individual spot checks with hand calculations.

Raw data included the number of days absent for each student. This variable was recoded into a binary determination of whether the student exhibited absenteeism in the 2008-2009 school year, as defined by London, Sanchez and Castrechini (2016), is missing ten percent or more of the school year - typically eighteen days. Students missing eighteen or more days were coded as being chronically absent.

The district had great language diversity in the original dataset, with four languages in the final sample: Spanish (n=817), Vietnamese (n=7), Turkish (n=1) and Korean (n=2). Students listing Spanish as their primary language were coded as 1 for the model. Ethnicity was coded into a separate variable for each response included in the dataset.

Initial Data Tests

In an ideal world, scientists would observe the same two outcomes for the same (or a perfectly equal) participant in order to truly test a hypothesis; however, since this is impossible, equal groups must be created. If groups are determined to be statistically equal, a causal effect can be estimated. Randomized grouping has long been the gold standard for creating comparable groups; however, this is often impossible for ethical reasons (a treatment that is known to have negative outcomes) and is more practical in a clinical trial setting. An alternative, commonly used approach is statistical matching.

The goal of the present study was to determine language outcomes seven years after grade retention for English learners. Because random sampling was not an option, propensity score matching (PSM) was utilized in order to create statistically matched groups post-hoc, mimicking the golden standard as closely as possible for the purpose of analyzing a treatment effect, before statistical matching.

The initial complete dataset contained English learners students who attended school within Oklahoma City Public Schools from the 2008-2009 to the 2014-2015 school years without gaps. A simple, direct comparison of the two groups would be a demonstration of the differences between retained and promoted students in the group; however, this would result in a comparison of apples to oranges, so to speak, as retained students would, surely, be characteristically different. Pearson Chi-square and independent samples t-test analyses were conducted for the entire sample in order to show the difference between each group (retained and not retained) for each covariate. There were four significant differences in baseline data between the groups. TABLE 5 shows the p values for each initial test; for the complete sample, grade, student's September 1 age compared to grade, absenteeism patterns, and listening score.

Table 5
Chi-Squared Goodness of Fit between Retained and Promoted Groups– Complete Sample

Covariates	**Treatment**	**Control**	X^2	**p**
Initial Grade	129	699	10.697	.005*
Special Education	6	54	1.531	.216
Age for Grade	119	486	28.567	.000*
Absenteeism	13	34	5.528	.019*
Language: Spanish	129	688	2.057	.151
Asian	0	9	1.679	.195
African American	1	2	.722	.396
Native American	1	3	.271	.603
White	1	4	.075	.785
Gender (Male)	49	324	3.08	.079
Free/Reduced Lunch	80	479	2.105	.147
Born Outside of US	116	623	.072	.789
School Reading	107	569	.173	.677

Table 6
Independent Samples T-test between Retained and Promoted Groups– Complete Sample

Covariates	N	*M*	*SD*	t	*p*
Listening PL	t=129 c=699	3.464 4.111	1.445 1.238	-5.313	.000
Speaking	t=129 c=699	3.008 3.995	1.585 1.625	-6.366	.077
Reading	t=129 c=699	2.423 3.288	1.313 1.374	-6.615	.282
Writing	t=129 c=699	1.966 2.494	.623 .615	-8.939	.434

It is intuitive that the groups would be different; students in the retained group are inherently different from students who were continuously promoted; the decision to retain might be school-based (such as an administrator who discouraged retention), but in many cases, students simply achieved enough that they were never at risk for retention. These differences justify the use of PSM to reduce biases in comparison groups so an effect of retention can be estimated.

Another initial test was conducted in order to check statistical differences between groups over time. A one-way repeated measures ANOVA was conducted to show changes in language scores over time for each group (retained and promoted). There was a significant effect for retained students on each language score. For Listening Scale Scores, Wilks' Lambda=.974, F (6, 821)=3.721, p= .001, Reading Scale Scores, Wilks' Lambda=.856, F (6, 821)=23.099, p= .000, and Writing Scale Scores, Wilks' Lambda= .159, F (6, 821)=27.242, p= .000. Plots for scores over time may be found in Appendix A.

Speaking was eliminated as a dependent variable due to the variability of growth and lack of normal distribution. Because speaking is scored manually by the test administrator, reliability error is suspected.

Table 7
Repeated Measures ANOVA

Effect	Wilks' Lambda	*Hypothesis df*	*Error df*	F	*p*	Partial η^2
Listening SS* Retained0910	.974	6	821	3.721	.000*	.026
Speaking SS* Retained0910	.910	6	821	13.534	.000*	.090
Reading SS* Retained0910	.856	6	821	23.099	.000*	.144
Writing SS* Retained0910	.834	6	821	27.242	.000*	.166

Propensity Score Matching

Grade retention studies began to utilize matched groups in the early 1990s. Reynolds (1992) matched retained and promoted students by considering age and basic demographic information. Jimerson (1999) selected matched promoted students for comparison based solely on being in the bottom quartile of achievement results. Other studies around the same time elected to use multiple regression instead of matching (McCoy & Reynolds, 1999; Rodney, Crafter, Rodney, & Mupier, 1999). Silberglitt, Jimerson, Burns and Appleton (2006) acknowledged the use of propensity matching, citing the work of Rosenbaum and Rubin (1983) as significant for other fields, but indicating that such matching was unnecessary for their study, which matched students on assessment results, finding a lack of significance among background variables between test groups. Wu, West and Hughes (2008) used propensity score matching in its modern form, matching students on 72 background variables, including demographic

variables, parent and teacher data on academic aptitude, academic achievement, personality indicators, peer relations, and family data. Additional recent studies (Im, Hughes, Kwok, Puckett, & Cerda, 2013; Raffaele Mendez, Kim & Ferron, 2015) have taken the same approach, with only small differences in variables.

PSM Model Design

The PSM procedure adjusts for selection bias by matching retained and promoted students according to background variables from the school year before students were retained (2008-2009 or 2009-2019). This process yielded two groups with like characteristics: one group (retained) included students who were retained in the 2008-09 or 2009-10 school years, and the other group included students of similar background traits who were continually promoted throughout the study timeframe.

Tumlison, Sass and Cano (2014) describe in detail the procedure for utilizing propensity score estimates in research. The first step is building the model, which includes a careful selection of covariates. Next, the researcher considers how to identify the best pairs with matching type, caliper and replacement. Once matches are established, the third step is to assess the balance and characteristics of the matches. After this, the data may be analyzed for treatment effect.

Covariates for Matching

Early propensity score matching typically utilized one or a small number of covariates to statistically match subjects (Rubin 1973a, Rubin 1973b, Cochran & Rubin, 1973); however, modern approaches include a much larger number. Rubin and Thomas (1996) recommended that all possible variables should be included unless there is evidence for excluding them, and Stuart (2010) recommends a liberal approach to inclusion of variables, asserting that there is little

chance of harming the model by including variables that end up being statistically unassociated with the effect.

The covariates were 22 background characteristics taken from the student information system in the 2008-2009 school year, before any students were retained. These independent variables were utilized to contribute to the prediction of the dependent variable for PSM, retention. The independent variables included initial grade, Special Education status, relative age (to grade), absenteeism (missing 18 or more days), home language (Spanish), language proficiency levels (listening, speaking, reading and writing) in 2008-2009 (before potential retention), gender (male), ethnicity (Hispanic), identifying as born in the US, free/reduced lunch status, and being in a school that met third grade reading benchmarks. SPSS was used to run the analysis with an alpha level of .05. Each variable has significance in the retention literature. Appendix B provides descriptive statistics for each PSM covariate.

Initial grade. Two grade-related variables were included. Initial grade was taken from the first year of the study, meaning the grade students were in for the 2008-2009 school year. Grade retention investigations have shown this to be a contributing variable for the likelihood of retention (Martin, 2011); for example, students are more likely to be retained in kindergarten or first grade (Warren & Saliba, 2012). Students were in kindergarten (n=291), first (n=310) or second grade (n=227) for the 2008-2009 school year. Students in kindergarten were coded as 1 for the model.

Age-grade comparison. Even more common than initial grade is some assessment of student age, which is very frequently included as a background variable in studies of a similar design (Hughes, West, Kim & Bauer, 2018; Hughes, Cao, West, Smith & Cerda, 2017, Mathys, Véronneau & Lecocq, 2017; Klapproth et al., 2016; Fruehwirth, Navarro & Takahashi, 2016).

This is likely based around the idea that students who are younger for the typical age are at a higher risk for repeating a grade (Burkam, LoGerfo, Ready & Lee, 2007). Oklahoma City Public Schools Board of Education Policy F-05 requires that students be five years of age on September 1 of the current school year in order to enter kindergarten (OKCPS, 2005). Students who were of the typical age for their grade (n=605) were coded as 1, as they are more likely to be retained than students who are older for their cohorts (n=223).

Special Education. Having a Special Education Individual Education Plan (IEP) is another common covariate in grade retention studies. Retained students have been found to be more likely to have an IEP either before retention (McLeskey, Lancaster & Gizzle, 1995) or after (Raffaele Mendez, Kim, Ferron & Woods, 2015; Roderick & Nagaoka, 2005). Students with an IEP during the 2008-2009 school year were coded as 1 (n=60).

Absenteeism. Chronically missing school for excused or unexcused reasons, is of growing concern in the academic literature, as it can be both a warning sign and the cause for a student to fall behind. Absenteeism, defined as missing 10% or more of the school year (typically 18 days) can be a major predictor of academics later on, including both achievement and graduation status (London, Sanchez & Castrechini, 2016), and is frequently used as a predictor (Rhodes, Thomas & Liles, 2018; Martin, 2011; Jimerson, Carlson, Rotert, Egeland, Sroufe, 1997). Students with 18 or more absences for the 2008-09 school year were coded as 1 for absenteeism (n=47).

Home language. The sample included only students qualifying as English learner, which is a only matching variable in other studies (Marsh, 2016; Goldstein, Eastwood & Behuniak, 2014; Wu, West & Hughes, 2008). Because the present study seeks to explore grade retention outcomes on language development, it was necessary to exclude native English speakers.

Because EL students are a very diverse subgroup, a number of language-related variables were used to secure more precise matches. First, home language was included as a covariate (Spanish=817, Vietnamese=7), as native language can affect academic achievement (Goos et al., 2013; Crosnoe & Turley, 2011).

Proficiency level. The student's language proficiency level was included as a covariate. The assessment administered to all English learners, WIDA, provides a scale of language development levels from 1.0 (beginner) to 6.0 (near-native speaker). Language proficiency are variables of great interest because reading is a primary reason for retention (Range, Pijanowski, Holt, & Young, 2012; Bowman, 2005; Tomchin & Impara, 1992; Rafoth & Carey, 1991). Gauging language of study participants is often central to study design (Fruehwirth, Navarro & Takahashi, 2016; Cham, Hughes, West & Im, 2015; Goldstein, Eastwood & Behuniak, 2014; Im, Hughes, Kwok, Puckett & Cerda, 2013; Anfinson, 1941).

It is important to note that a separate WIDA score, the scale score (as opposed to the proficiency level) is utilized as a dependent variable in the final model. Scale scores are vertically aligned so they are comparable across grade bands, making them an appropriate measure of language growth for students in different grades; they are also the optimal way to gauge year to year progress for a single student (Kenyon, MacGregor & Cook, 2011).

Country of origin. Some English learners were also born outside of the United States; estimates suggest that about 75% of children in newly-immigrated families are born in the Unites States (Fix & Zimmerman, 2001); however, citizenship numbers are not tracked in schools, as it could be a deterrent to enrollment (Plyler v. Doe, 1982). Still, being an immigrant has been shown to affect academic outcomes (Tillman, Guo & Harris, 2006) and increase a student's chances of being retained (Marsh, 2016; Goos et al., 2013). In many states, including Oklahoma,

parents are asked to identify a place of birth, but it does not require verification. In the present study, students identified by a parent or guardian as having been born outside of the United States were coded as 1 (n=89).

Ethnicity. Demographic characteristics are another common variable for matching, with ethnicity appearing as a variable of significance in a number of retention studies (Fruehwirth, Navarro & Takahashi, 2016; Cham, Hughes, West & Im, 2015; Chen, Hughes & Kwok, 2014; Demanet & Van Houtte, 2013). Home language is the language listed by parents (or students of age) upon completion of enrollment documents. Primary language of the household or preferred test language are often utilized as matching variables to determine student language background (Goos, Van Damme, Onghena, Petry & de Bilde, 2013; Im, Hughes, Kwok, Puckett & Cerda, 2013).

Sex. Males are more often targeted for retention (Tingle, Schoeneberger & Algozzine, 2012; Malone, West, Denton, Park, National Center for Education Statistics & Education Statistics Services Inst., 2006; El-Hassan, 1998; Jimerson, Carlson, Rotert, Egeland & Sroufe, 1997; Byrd & Weitzman, 1994; Abidin, Golladay & Howerton, 1972). Many studies have utilized gender as a variable in similar models (Fruehwirth, Navarro & Takahashi, 2016; LiCalsi, Ozek & Figlio, 2016; Marsh, 2016; Chen, Hughes & Kwok, 2014). In the present study, males are coded as 1 (n= 455).

Socioeconomic status. A student's free/reduced price lunch status is a common measurement of the socioeconomic status of his/her family. Socioeconomic status is one of the most significant predictors of a student's future academic success (Kim, Mazza, Zwanziger & Henry, 2014; Van Ewijk & Sleegers, 2010; McGill-Franzen, Zmach, Solic & Zeig, 2006). Additionally, low socioeconomic status (SES) is a common predictor of being retained (Dong,

2010; Malone, West, Denton, Park & NCES, 2006); in fact, it has been asserted that children of poverty are more likely to be retained because their parents lack the social capital that comes with middle class status (LiCalsi, Ozek & Figlio, 2016; Stearns, 2007; Akmal & Larsen, 2004). For these reasons, low SES is used as a matching variable frequently in retention literature (Marsh, 2016; Fortner, & Jenkins, 2016; LiCalsi, Ozek & Figlio, 2016; Fruehwirth, Navarro & Takahashi, 2016; Goldstein, Eastwood & Behuniak, 2014; Chen, Hughes & Kwok, 2014; Demanet & Van Houtte, 2013; Stearns, 2007).

School reading scores. One school-related covariate was included, as school achievement has sometimes been shown to influence individual student success (Feld & Zolitz, 2017; Antecol, Eren & Ozbeklik, 2016; Burke & Sass, 2013). School report cards were used (Office of Educational Quality & Accountability, 2010). The variable was coded as 1 if the student's school met the state benchmark for third grade reading (n=152).

Specifications for Determining Propensity Score

A few specifications are decided before a propensity score can be estimated. Tumlinson, Sass and Cano (2014) describe in detail the decisions that must be made in crafting the right model. These include matching style, caliper and ratio matching.

The first step the propensity matching design is to determine a matching style. The most direct style is exact matching, where only cases with exactly the same propensity scores are matched; however, this often results in a loss of many subjects (Rosenbaum & Rubin, 1985). More common are styles of "greedy" or "local" matching, which incorporate some flexibility (Stuart, 2010). Nearest neighbor matching allows the model to select one subject from the treatment groups and find a close match in the control group (Cochran & Rubin, 1973). Matching can also be done with replacement using a ratio, n:1 (Ming & Rosenbaum, 2000),

which allows one treatment subject to be matched to a predetermined number of controls. Allowing a variable number of controls has been found to reduce bias in matched-group studies (Ming & Rosenbaum, 2000). The present study will utilize 5:1 matching, allowing up to five controls to be matched to one treatment, as recommended by Ming and Rosenbaum (2000).

Caliper is the distance between matches as measured in standard deviations from the mean (Rosenbaum & Rubin, 1985a; Cochran & Rubin, 1973). There is no agreed upon caliper distance in the literature (Austin, 2011); instead, researchers must find a caliper that best fits the data (Tumlinson, Sass & Cano, 2014). A higher caliper may be set, but at risk of low-quality matches; in contrast, lower calipers create better quality matches, but there may be fewer matches to include in the study (Austin, 2011; Cochran & Rubin, 1973). This becomes a balance between finding as many matches as possible and using only pairs that are very similar to one another (Rosenbaum & Rubin, 1985). A generally accepted caliper is 0.25 (Tumlinson, Sass & Cano, 2014); this a common caliper retention literature as well (West et al., 2014; Hughes, Chen, Thoemmes & Kwok, 2010; Wu, West & Hughes, 2008). Fit tests of the PSM model to assess caliper fit and the balance of the matches will be described in the next chapter with the results.

The final model was run using the PS Matching R plug in for SPSS (Thoemmes, 2012). With the above specifications, and utilizing logistic regression, propensity scores were estimated to identify matched groups (Tumlinson, Sass, & Cano, 2014). In order to control for the differences and error in these matches across treatment and control groups, or control for the covariates which account for the differences in selection, PS Matching creates a sampling weight for each case. This weight is used in the final model to test the research questions.

Data Analysis

Measuring the Research Question: Language Growth

Though state reading and math testing begins in third grade, language is tested annually starting in kindergarten, with vertically aligned WIDA ACCESS scores (Kenyon, MacGregor, Li & Cook, 2011), which can be used to assess language growth from year to year. Students are assessed in four domains: Listening, Speaking, Reading and Writing. Three of the domains are utilized as dependent variables (Listening, Reading and Writing). Due to the longitudinal design of the study, a multilevel growth model can demonstrate the influence of predictors (such as treatment/control groups) on both the growth trajectories of students over time, plus the starting point, or intercept, for achievement.

Data Restructure and Cleaning

Data were converted from a wide form to a long form for analysis in SPSS. The original dataset contained one case per subject, which reflected Listening, Reading and Writing scores for each of the seven years. The restructured dataset reflects one line per test, per year (3 tests x 7 years) and a new variable indicating the year of the study (1, 2, 3, etc.). This restructuring of data allows for the creation and coding of time variables.

An assessment of growth trends indicated the need for both linear and quadratic trajectories. A new variable, *quadtime*, was created for incorporation into the model. See Appendix C for growth trend dot plots. Additionally, in order to decrease multicollinearity of data, it is recommended to use orthogonal time components, which creates a grand mean of zero for the data, minimizing correlation between time-related variables (Heck & Thomas, 2015; Heck, Tabata & Thomas, 2013). This is particularly suitable for time data which are evenly space, and often help to stabilize model estimation. An additional variable, *orthotime*, was added

to the dataset. Orthogonal polynomials are set in a way that each number is at a right angle to its counterpart. In the current model, linear time across seven years is a simple sequence, starting with zero (0, 1, 2, 3, 4, 5, 6). The orthogonal conversion hinges on the midpoint, year four, represented by the number 3 (third year of the study), creating a different sequence that is less likely to cause interaction with time points (-3, -2, -1, 0, 1, 2, 3).

Though utilized in the PSM model, a number of variables were excluded from the final model. Ethnicity (Native American, African American, white) and Language (Spanish) holdout groups were removed from the model due to complete or near-homogeneity. Speaking and Reading proficiency levels were also removed due to strong correlations among the four language domains (Listening, Speaking, Reading and Writing). Two domains with the least correlation were retained in the model: Listening represented both verbal and receptive language, and Writing represented literacy and expressive language. Further, to add to interpretation, all scale variables used as predictors were grand-mean centered according to the raw, unweighted data (Heck, Tabata & Thomas, 2013).

Final Model

Data were analyzed in MPlus using a multilevel model to examine change in individual growth trajectories over time at level 1, and changes between individuals at level 2 (Heck & Thomas, 2015). The multilevel model is superior to a simpler model, such as a repeated-measures ANOVA, because the model can analyze time measurements nested within the individual which adjusts the standard error. The model also utilizes the same algebraic form for each change trajectory, but can analyze differences in slope and intercept for each individual. The model utilizes an unstructured, symmetric covariance matrix with intercept and slope in the diagonals. The Level 1 equation is as follows:

Level 1: $Y_{it} = \pi_{oi} + \pi_1 growthrate_{it} + \varepsilon_{it}$

The equation utilizes Y_{ti} to indicate the time points for the dependent variable, π_{01} to represent the intercept for initial status, π_{1i} to represent expected linear growth rate for the individual over time, and ε_{it} to represent residuals, or random error.

At Level 2, β_{x0} represents the intercept and, β_{xi} fixed effect for each individual-level predictor.

$\Pi_{0i} = \beta_{00} + r_{0i}$

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

Chapter 4: Results

Introduction

The present study seeks to discover outcomes for English learner students over a seven-year period to determine outcomes of retained students as compared to matched-promoted peers. This chapter describes the results of two major steps: propensity score estimation and matching to create comparable groups, and longitudinal analysis with a multilevel growth model in order to estimate the impact of these groups on outcomes (slope and intercept) over time.

Propensity Score Matching

The process of estimating propensity scores and determining optimal matches is data-dependent (Tumlison, Sass & Cano, 2014) and made stronger by a robust set of background variables (Rubin, 2007). A typical PSM procedure requires a number of steps, described by Stuart (2010) and requires defining closeness, selection of a matching method, and assessment of matching results. These procedures are described in detail as the first part of data analysis since they determine the validity of the tested impact in the final model.

Propensity Score Models

Caliper must be a balance between finding the most matches and finding quality matches (Rosenbaum & Rubin, 1985a, Rosenbaum & Rubin, 1985b). The exact caliper should be set based on the fit of the data; that said, 0.25 is a common caliper in retention research (West et al., 2014; Hughes, Chen, Thoemmes & Kwok, 2010; Wu, West & Hughes, 2008). For this reason, the matching was done with caliper set a 1.0, .5 and 0.25. Each model generated a new dataset with a new PS Matching variable to indicate inclusion in the final model.

The PSM process is designed to create comparable groups that are statistically similar; for this reason, the next step in a PSM approach is to check that the model is a good fit for the

covariates included (West et al., 2014; Tumlinson, Sass & Cano, 2014; Thoemmes, 2012; Stuart 2010). Before matching, the differences between treatment and control groups were statistically significant on four of the background covariates: initial grade, student's September 1 age compared to grade, absenteeism patterns, and listening score. Two sets of tests were conducted on the covariates of the new sample in order to determine balance. First, t-test (string variables) and chi square (categorical variables) tests checked for significant differences between groups for each covariate in isolation. Table 8 reports the significance for each variable; no significant difference was found on individual covariates, meaning that the groups were well-balanced. The model with a caliper of 0.25 was retained.

A second test assesses the balance of the model through use of a logistic regression (Hansen & Bowers, 2008), simultaneously assessing balance for all covariates. The regression indicated only one significant difference: initial grade (p=0.29). A detailed comparison of *p* values for each covariate may be found in Table 9.

Table 8
Significance of t-test and chi square tests for Varying Caliper Levels

	Complete Dataset	**Caliper 1.0**	**Caliper .5**	**Caliper .25**	**Caliper .25 Weighted**
Sample	t=129 c=699	t=115 c=371	t=100 c=309	t=93 c=271	t=93 c=271
Initial Grade	.05*	.159	.420	.963	.924
SpEd	.216	.564	.944	.754	.887
AgeforGrade	.000*	.001*	.041*	.114	.523
Absenteeism	.019*	.044	.438	.335	.117
Language: Spanish	.151	.264	n/a	n/a	n/a
Asian	.195	.264	n/a	n/a	n/a
Black	.396	.379	.569	.557	.371
Native Am	.603	.691	.719	.427	.277
White	.785	.691	.399	.557	.248
Gender: Male	.079	.138	.899	.263	.433
F/R Lunch	.147	.211	.664	.570	.702
Born Outside US	.789	.879	.745	.777	.538
School Met Rdg.	.677	.829	.92	.905	.627
L	.000*	.005*	.010*	.835	.439
S	.077	.203	.629	.793	.660
R	.282	.533	.624	.241	.502
W	.434	.883	.779	.421	.536
Significant Tests	4	2	2	0	0

Table 9
Assessing Covariate Balance with Logistic Regression

	Unstandardized Coefficients		**Standardized**		
	B	**Std. Error**	**Beta**	**t**	**Sig.**
(Constant)	.407	.133		3.057	.002
Initial Grade	.100	.046	.178	2.189	.029
Special Ed.	-.081	.100	-.044	-.813	.417
Age for Grade	.120	.064	.101	1.869	.062
Absenteeism	.064	.088	.039	.730	.466
African American	-.331	.444	-.040	-.745	.457
Native American	.256	.312	.043	.820	.413
White	-.352	.442	-.042	-.797	.426
Gender (Male)	.053	.047	.060	1.111	.267
Free/ Reduced	-.036	.049	-.040	-.746	.456
Born Outside US	-.014	.082	-.010	-.177	.859
Listening PL	.008	.021	.023	.365	.715
Speaking PL	-.020	.018	-.073	-1.119	.264
Reading PL	-.028	.028	-.086	-.980	.328
Writing PL	-.104	.066	-.142	-1.566	.118
School Met Rdg	.039	.061	.035	.641	.522

Propensity scores generated by the model predict the likelihood that an individual will be assigned to the treatment group (retention) based on background variables that have been found to be relevant according to the literature. The PSM yielded a number of treatment (n=93) and control (n=271) participants to be included for further analysis.

After propensity scores were estimated, a test between groups, called the common support area, is a check to ensure that there is overlap among the retained and promoted matched groups. A probability score between 0 and 1 is indicated for each student, predicting the likelihood for the individual to be retained; a score of 0.0 would indicate no chance of retention, whereas a score of 1.0 would indicate a 100% chance of grade retention. According to West et al. (2014), causal effects may only be estimated for data included in the common support region. The range of common support was a high, with a minimum of .005 in the retained group, and a maximum of .865 in the promoted group, which allowed 364 cases to be included in the model (see Table 10). Cases outside of the common support area were discarded. The wide range of propensity scores also supports the use of the chosen matching model.

Table 10
Descriptive Statistics for Propensity Scores, By Group

	N	Minimum	Maximum	Mean	St. Dev.
Retained	93	.005	.913	.35863	.227544
Matched Promoted	271	.000	.865	.21305	.174240

PS Weights

Propensity score matching is used to select cases for inclusion in a balanced model. Each case is given a PS (propensity score) weight, which is typically included for model analysis (Stuart, 2010). This accounts for ratio matching, ensuring equality of representation among otherwise unbalanced numbers of control and treatment cases (Thoemmes, 2012). Table 12 shows descriptive statistics for the data before case weights were applied, and then after. The final model utilized the weighted data.

Table 11
Descriptive Statistics for Covariates Before and After PSM

	Full Sample (n=828)				Analytic Sample (n=364)			
	Min	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**
Initial Grade	.00	1.00	.3514	.47771	.00	1.00	.3929	.48906
SpEd	0	1	.07	.259	0	1	.06	.239
Age – Avg.	.00	1.00	.7307	.44388	.00	1.00	.8407	.36650
(Age is +1)	.00	1.00	.2536	.43535	.00	1.00	.1566	.36392
(Age is +2)	.00	1.00	.0133	.11456	.00	1.00	.0027	.05241
Absenteeism	.00	1.00	.0568	.23153	.00	1.00	.0742	.26242
Spanish	.00	1.00	.9867	.11456	1.00	1.00	1.0000	.00000
(Vietnamese)	.00	1.00	.0085	.09161	.00	.00	.0000	.00000
(Hispanic)	.00	1.00	.9746	.15732	.00	1.00	.9890	.10439
Asian	.00	1.00	.0109	.10375	.00	.00	.0000	.00000
Black	.00	1.00	.0036	.06012	.00	1.00	.0027	.05241
Native Am	.00	1.00	.0048	.06938	.00	1.00	.0055	.07402
White	.00	1.00	.0060	.07752	.00	1.00	.0027	.05241
Male	.00	1.00	.5495	.49784	.00	1.00	.5742	.49515
(Female)	.00	1.00	.4505	.49784	.00	1.00	.9890	.10439
F/R Lunch	.00	1.00	.6751	.46861	.00	1.00	.6374	.48142
ID as Born Outside US	.00	1.00	.1075	.30992	.00	1.00	.0934	.29140
School Rdg	.00	1.00	.1836	.38737	.00	1.00	.1978	.39889
L	1.0	6.0	4.010	1.2930	1.2	6.0	3.914	1.3115
S	1.0	6.0	3.842	1.6573	1.0	6.0	3.582	1.6071
R	1.0	6.0	3.153	1.3990	1.0	5.5	2.884	1.3616
W	1.0	4.4	2.411	.6449	1.0	3.4	2.195	.5949

Table 12
Descriptive Statistics for Covariates Before and After PSM, Plus Weights

	Full Sample (n=828)		Analytic Sample (n=364)			
			Unweighted		Weighted	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Initial Grade	.9227	.78764	.8434	.77863	.4091	.49234
SpEd	.07	.259	.06	.239	.05	.220
Age – Avg.	.7307	.44388	.8407	.36650	.8734	.33299
(Age is +1)	.2536	.43535	.1566	.36392	.1239	.32987
(Age is +2)	.0133	.11456	.0027	.05241	.0027	.05241
Absenteeism	.0568	.23153	.0742	.26242	.0777	.26806
Spanish	.9867	.11456	1.0000	.00000	1.0000	.00000
(Vietnamese)	.0085	.09161	.0000	.00000	.0000	.00000
(Hispanic)	.9746	.15732	.9890	.10439	.9890	.10454
Asian	.0109	.10375	.0000	.00000	.0000	.00000
Black	.0036	.06012	.0027	.05241	.0016	.04004
Native Am	.0048	.06938	.0055	.07402	.0067	.08199
White	.0060	.07752	.0027	.05241	.0027	.05166
Male	.5495	.49784	.5742	.49515	.6072	.48904
(Female)	.4505	.49784	.9890	.10439	.3928	.48904
F/R Lunch	.6751	.46861	.6374	.48142	.6214	.48570
Born Outside US	.1075	.30992	.0934	.29140	.0785	.26941
School Rdg	.1836	.38737	.1978	.39889	.1850	.38884
L	4.010	1.2930	3.914	1.3115	3.806	1.3538
S	3.842	1.6573	3.582	1.6071	3.445	1.6156
R	3.153	1.3990	2.884	1.3616	2.715	1.3343
W	2.411	.6449	2.195	.5949	2.113	.5892

Longitudinal Analysis

A multilevel growth model was utilized to analyze the effects of student characteristics, including retained (treatment) versus promoted (matched control), on language outcomes over seven years. The presentation of results below reflects the guidelines by Heck and Thomas (2015). Models were run separately for each dependent variable: listening, reading and writing. A multilevel growth model follows a stepwise process in which the first model, null or empty, assesses the amount of variance to explain in the dependent variables, next individual level

predictors are added in a second model. The longitudinal analysis below did not have time-varying covariates.

Null: Model 1

Fit indices for each dependent variable are presented in Table 13, confirming six estimated parameters for each model. The coefficient estimates for listening, reading and writing are presented in Tables 15, 16 and 17. For the model initial status, the coefficient is 338.096 for Listening, 315.511 for Reading, and 308.871 for Writing, representing the Year 4 intercept. The growth rate model indicates a coefficient of 32.499 for Listening, for 17.213 Reading and for 31.937 Writing, indicating average growth per year for each domain.

The percentage of the variance to be explained in the model without predictors is: 12.7% for Listening, 12.3% for Reading, and 12.9% for Writing, which is large enough to continue the analysis. For initial status variance, the z-test coefficient is significant for each ($p < 0.001$): Listening is 10.785, Reading is 10.862, and Writing is 11.865, indicating that there is significant variation between individuals in the model. Similarly, the z-test coefficient was significant for each domain ($p < .001$): the coefficient for Listening was 6.288, Reading was 11.252, and Writing was 11.213, indicating that there is significant variance in growth rates. With enough variance to explain across years by individual students, predictors were added into the model.

Table 13
Model 1 Fit Indices

	Listening	Reading	Writing
Number of Free Parameters	6	6	6
Log Likelihood			
H_0 Value	-12121.360	-12273.428	-12076.538
Information Criteria			
Akaike (AIC)	24254.719	24558.856	24165.076
Bayesian (BIC)	24289.778	24593.914	24200.134
Sample-Size Adjusted BIC ($n^* = (n + 2) / 24$)	24270.714	24574.850	24181.071

Table 14
Variance for Model 1

Listening	Reading	Writing
12.7%	12.3%	12.9%

Table 15
Model 1 (Null) Unconditional Model of Linear Growth – Listening

		Standard		
	Coefficient	Error	Est./SE	P-Value
Fixed Effects				
Mean initial status, $\boldsymbol{\beta}_{00}$	338.096	0.984	343.486	0.000
Mean growth rate, $\boldsymbol{\beta}_{10}$	32.499	0.380	42.001	0.000
Random Effects				
Initial status, u_{0i}	272.646	25.281	10.785	0.000
Growth rate, u_{1i}	32.499	5.168	6.288	0.000
Covariance	-11.491	8.986	-1.27	0.201
Level-1, ε_{ti}	559.830	63.495	8.817	0.000

Table 16
Model 1 (Null) Unconditional Model of Linear Growth – Reading

		Standard		
	Coefficient	Error	Est./SE	P-Value
Fixed Effects				
Mean initial status, $\boldsymbol{\beta}_{00}$	315.511	1.075	293.428	0.000
Mean growth rate, $\boldsymbol{\beta}_{10}$	17.213	0.445	38.658	0.000
Random Effects				
Initial status, u_{0i}	325.829	29.997	10.862	0.000
Growth rate, u_{1i}	47.966	4.263	11.252	0.000
Covariance	-115.532	10.139	-11.393	0.000
Level-1, ε_{ti}	672.066	44.571	15.079	0.000

Table 17
Model 1 (Null) Unconditional Model of Linear Growth – Writing

	Coefficient	Standard Error	Est./SE	P-Value
Fixed Effects				
Mean initial status, β_{00}	308.871	1.120	275.660	0.000
Mean growth rate, β_{10}	20.522	0.379	54.206	0.000
Random Effects				
Initial status, u_{0i}	376.134	31.702	11.865	0.000
Growth rate, u_{1i}	31.937	2.848	11.213	0.000
Covariance	100.608	8.302	-12.119	0.000
Level-1, ε_{ti}	565.291	27.415	20.620	0.000

Individual-level Predictors: Model 2

Model 2 incorporates covariates in order to explain the variance in the outcome variables across time. As such, the fit indices for each of the three outcome variables are presented in Table 18.

Table 18
Model 2 Fit Indices

	Listening	Reading	Writing
Number of Free Parameters	34	34	34
Log Likelihood			
H_0 Value	-11764.410	-11923.545	-11677.925
Information Criteria			
Akaike (AIC)	23596.819	23915.090	23423.850
Bayesian (BIC)	23795.483	24113.755	23622.514
Sample-Size Adjusted BIC ($n^* = (n + 2) / 24$)	23687.456	24005.727	23514.487

Tables 19, 20 and 21 show the model results for the set of predictors. The coefficients for these predictors demonstrate similar findings for each of the three outcome variables. These predictors were regressed on the initial status of the outcomes variable, which was Year 4, as well as on the slopes which is the growth rate.

Influence of Retention, Age and Grade

Students generally made progress in their language development, growing more proficient each year of the study (see Appendix A). This is expected, as students were learning in the target language, meaning that they would have listened, spoken, read and written in English as a part of their time in school. As a whole, language development growth is often faster in the beginning, and slows as students attain higher proficiency (Cook, Boals & Lundberg, 2011.

Retention status, grade and relative age predicted student achievement in Year 4 for both slope and intercept. Grade retention intercept was significant for all three dependent variables. Scale scores for retained students were significantly lower for Listening (-8.172, $p < .001$), Reading (-11.517, $p < .001$) and Writing (-11.495, $p < .001$). There was no significant difference in growth rates of retained compared to promoted students for Listening, Reading or Writing Scale Score outcomes. While retained and promoted students did not differ in their growth rates, we know from the results above that the retained students performed lower and stay lower since their growth rate is parallel to those who were promoted. There was no opportunity provided for English learners to catch up. The sorting of students into lower grades and overall concurrent language performance were the main indicators of static ability in Year 4, midway through the seven-year study.

Students with a lower initial grade predicted a decreased WIDA scale score, but a higher growth trajectory. Students whose initial grade was kindergarten had significantly lower scale score intercept in Listening (-15.81, $p < .001$), Reading (-19.969, $p < .001$) and Writing (-12.765, $p < .001$) in Year 4. In contrast, growth rates for students starting in kindergarten were higher: Listening (3.946, $p < .001$), Reading (9.488, $p < .001$), and Writing (4.733, $p < .001$) indicated

that students retained in an earlier grade (kindergarten versus first or second) had an increased growth rate.

Student age was measured as expected for grade, as compared to district policy for target school entrance age. Because younger students are more likely to be retained (Burkam, LoGerfo, Ready & Lee, 2007), students of average age for grade, as opposed to older for cohort, were examined. Being of average age for grade was significant for intercept and some slope measurements. Students who were of the expected age for their grade had higher intercepts for Listening (7.040, $p < .001$), Reading (4.279, $p < .005$) and Writing (4.693, $p < .01$). Students of expected age for cohort indicated higher scores than their older-for-cohort peers. Being of average age for grade was also significant in growth trajectory only for Reading (1.841, $p < .01$) and Writing (1.084, $p < .05$).

Influence of Individual Characteristics

Four individual student characteristics were examined to determine the influence of background characteristics on student achievement. Beyond grade and age, other student and school factors influenced particular areas of language performance.

Students with free or reduced lunch exhibited no differences for slope on any dependent variable, but had higher Reading intercept (2.646, $p < .05$); this effect was small but significant, and is inconsistent with modern literature and commonly-disseminated information that children of poverty struggle in reading (McGill-Franzen, Zmach, Solic & Zeig, 2006), supported by the oft-cited 30 million word vocabulary gap (Hart & Risley, 1992). However, modern attempts to replicate the famous study with larger populations have found that a vocabulary gap, if it exists for children of low socioeconomic status, was likely overstated in prior literature, and that the one consistency is inconsistency among families (Sperry, Sperry & Miller, 2018; Gilkerson,

2017). In short, the small reading increase for children of poverty is not inconsistent with modern literature, indicating that there is variance among families of poverty, not a pronounced language deficit.

Students with a pattern of absenteeism indicated a small but significant outcome for Reading Scale Score growth outcomes only (1.918, $p < .05$), indicating that students who were frequently absent had slightly higher reading growth. While this might seem counterintuitive, absenteeism was measured in the first year of the study, and may have improved in subsequent years, indicating more growth. Scores for male students were lower for Writing only in both slope (-1.262, $p < .001$) and intercept (-3.758, $p < .01$). Like above, the significance of sex on writing demonstrates that there are nuances based on gender within the population of English learners. Similarly, being born outside of the United States influenced Year 4 Writing. For intercept, students scored lower if they were identified as born outside of the United States (-3.891, $p < .05$). Slope was almost significant for Writing (-1.2016, $p < .1$). These findings demonstrate that the English learner designation intersects with other background characteristics and labels in schools.

Influence of Baseline Measures

Baseline student achievement indicators were also significant predictors. Students receiving Special Education services had lower listening scale scores (-8.126, $p < .01$), but intercept differences were not significant for Reading and Writing. Significant differences in slope emerged for all three dependent variables: Listening (-4.230, $p < .01$), Reading (-3.202, $p < .001$), and Writing (-1.922, $p < .01$). Special Education status influenced student outcomes, particularly slope, for dually identified learners.

Initial Listening and Writing Proficiency Levels were significant across the three dependent variables. Listening Proficiency Level (PL) was positive for intercept for all dependent variables: Listening (5.777, $p < .001$), Reading (2.359, $p < .001$) and Writing (1.185 ($p < .005$). In contrast, slope for Listening PL was negative for each dependent variable: Listening (-3.363, $p < .001$), Reading (-0.945, $p < .001$), and Writing (-0.339, $p < .005$). Initial Writing PL was a significant predictor for most, but not all, outcomes. As with Listening, Writing PL was positive for intercept for all three dependent variables: Listening (4.733, $p < .005$), Reading (10.920, $p < .001$) and Writing (18.618, $p < .001$). Writing PL was a strong predictor of Reading and Writing scale score intercept. Writing PL was also a significant predictor of slope for Reading (-3.666, $p < .001$) and Writing (-7.008, $p < .001$), but not Listening.

These significant characteristics in school achievement prior to retention demonstrate the importance of understanding more nuanced subpopulations within the English learner group, and indicate that initial achievement influence language outcomes years later.

Influence of School Context

A number of school variables were included at the student level. Two variables, percentage of Hispanic population and teacher experience, yielded no significant results for the dependent variables. The same was true for the Listening dependent variable, regarding whether or not the school met their third grade reading benchmark. However, students attending a school that met third grade reading targets had lower scale score outcomes in reading intercept (-3.707, $p < .05$) and slightly higher growth rates in writing growth rates (1.333, $p < .01$). These results show that aspects of the school influence how English language students perform; the surrounding school context provides either support or barriers for how English learners are able

to access skills. There are likely other strong factors at the school level that influence students' language development, not measured in this work, but which show themselves in higher dropout rates for English learners (Sheng, 2011) and Hispanic/Latino students (NCES, 2014; Stearns, 2007).

Table 19
Model 2 Linear Model of Language Development Growth – Listening

	Coefficient	Standard Error	Est./SE	P-Value
Fixed Effects				
Model for **initial status**, π_{0i}				
Intercept, β_{00}	340.126	2.887	117.809	0.000
Retained 09/10	-8.172	1.669	-4.895	0.000
Initial Grade	-15.810	2.171	-7.282	0.000
SpEd	-8.126	3.048	-2.666	0.008
AgeAverage	7.040	1.899	3.707	0.000
Absent	-0.814	2.569	-0.317	0.751
Gender-M	-1.631	1.289	-1.266	0.206
FR Lunch	2.348	1.363	1.723	0.085
BornOutsideUS	-0.952	1.753	-0.543	0.587
ListeningPL	5.777	0.498	11.607	0.000
WritingPL	4.733	1.974	2.398	0.016
SchRBenchmark	-1.745	1.543	-1.131	0.258
SchHispanic	1.042	1.711	0.609	0.543
Mobility	0.071	0.104	0.680	0.497
TeacherExp	0.076	0.220	0.346	0.729
Model for growth rate, π_{1i}				
Intercept, β_{00}	13.461	1.181	11.393	0.000
Retained 09/10	-0.327	0.637	-0.514	0.607
Initial Grade	3.946	0.929	4.248	0.000
SpEd	-4.230	1.577	-2.683	0.007
AgeAverage	1.160	0.807	1.437	0.151
Absent	-0.153	0.877	-0.175	0.861
Gender-M	-0.535	0.570	-0.938	0.348
FR Lunch	0.970	0.645	1.503	0.133
BornOutsideUS	0.624	0.698	0.894	0.371
ListeningPL	-3.363	0.208	-16.181	0.000
WritingPL	1.419	0.754	1.883	0.060
SchRBenchmark	-0.354	0.623	-0.568	0.570
SchHispanic	-0.007	0.705	-0.010	0.992
Mobility	0.030	0.043	0.690	0.490
TeacherExp	0.159	0.093	1.701	0.089
Random Effects				
Initial status, r_{0i}	74.732	17.591	4.248	0.000
Growth rate, r_{1i}	12.843	5.602	2.293	0.022
Level-1, e_{ti}	520.332	55.153	9.434	0.000
Covariance	30.791	6.792	4.534	0.000

Table 20
Model 2 Linear Model of Language Development Growth – Reading

	Coefficient	Standard Error	Est./SE	P-Value
Fixed Effects				
Model for **initial status**, $\boldsymbol{\pi}_{0i}$				
Intercept, $\boldsymbol{\beta}_{00}$	321.103	2.456	130.743	0.000
Retained 09/10	-11.517	1.438	-8.010	0.000
Initial Grade	-19.969	2.140	-9.331	0.000
SpEd	-2.443	2.761	-0.885	0.376
AgeAverage	4.279	1.523	2.809	0.005
Absent	-2.215	2.395	-0.925	0.355
Gender-M	-1.271	1.191	-1.067	0.286
FR Lunch	2.646	1.279	2.069	0.039
BornOutsideUS	0.057	1.558	0.037	0.971
ListeningPL	2.359	0.469	5.026	0.000
WritingPL	10.920	1.752	6.235	0.000
SchRBenchmark	-3.707	1.546	-2.397	0.017
SchHispanic	2.077	1.599	1.299	0.194
Mobility	0.063	0.090	0.697	0.486
TeacherExp	-0.163	0.208	-0.784	0.433
Model for growth rate, $\boldsymbol{\pi}_{1i}$				
Intercept, $\boldsymbol{\beta}_{00}$	11.807	1.018	11.601	0.000
Retained 09/10	0.311	0.598	0.520	0.603
Initial Grade	9.488	0.945	10.044	0.000
SpEd	-3.202	0.876	-3.656	0.000
AgeAverage	1.841	0.664	2.772	0.006
Absent	1.918	0.898	2.137	0.033
Gender-M	-0.673	0.501	-1.343	0.179
FR Lunch	1.044	0.538	1.938	0.053
BornOutsideUS	0.049	0.681	0.071	0.943
ListeningPL	-0.945	0.223	-4.246	0.000
WritingPL	-3.666	0.762	-4.808	0.000
SchRBenchmark	0.071	0.592	0.120	0.904
SchHispanic	-0.308	0.634	-0.486	0.627
Mobility	0.022	0.035	0.634	0.526
TeacherExp	-0.090	0.080	-1.130	0.259
Random Effects				
Initial status, r_{0i}	24.224	11.030	2.196	0.028
Growth rate, r_{1i}	0.339	2.801	0.121	0.904
Covariance	-1.149	3.830	-0.300	0.764
Level-1, e_{ti}	656.324	42.539	15.429	0.000

Table 21
Model 2 Linear Model of Language Development Growth – Writing

	Coefficient	Standard Error	Est./SE	P-Value
Fixed Effects				
Model for **initial status**, π_{0i}				
Intercept, β_{00}	313.038	2.575	121.569	0.000
Retained 09/10	-11.495	1.544	-7.444	0.000
Initial Grade	-12.765	2.276	-5.608	0.000
SpEd	-4.634	3.314	-1.398	0.162
AgeAverage	4.693	1.829	2.566	0.010
Absent	-0.744	2.453	-0.303	0.762
Gender-M	-3.758	1.196	-3.144	0.002
FR Lunch	1.850	1.309	1.413	0.158
BornOutsideUS	-3.891	1.953	-1.992	0.046
ListeningPL	1.185	0.504	2.349	0.019
WritingPL	18.618	1.918	9.706	0.000
SchRBenchmark	-0.768	1.545	-0.497	0.619
SchHispanic	2.110	1.576	1.339	0.180
Mobility	0.087	0.146	0.594	0.552
TeacherExp	-0.090	0.215	-0.420	0.675
Model for growth rate, π_{1i}				
Intercept, β_{00}	18.049	0.730	24.717	0.000
Retained 09/10	0.280	0.459	0.609	0.543
Initial Grade	4.733	0.600	7.891	0.000
SpEd	-1.922	0.575	-3.344	0.001
AgeAverage	1.084	0.480	2.261	0.024
Absent	0.252	0.714	0.353	0.724
Gender-M	-1.262	0.355	-3.557	0.000
FR Lunch	-0.033	0.356	-0.094	0.925
BornOutsideUS	-1.216	0.624	-1.950	0.051
ListeningPL	-0.339	0.142	-2.376	0.017
WritingPL	-7.008	0.532	-13.174	0.000
SchRBenchmark	1.333	0.409	3.255	0.001
SchHispanic	0.276	0.497	0.556	0.578
Mobility	0.042	0.039	1.075	0.283
TeacherExp	-0.061	0.059	-1.039	0.299
Random Effects				
Initial status, r_{0i}	53.716	10.415	5.157	0.000
Growth rate, r_{1i}	0.304	12.616	0.024	0.981
Level-1, e_{ti}	517.555	23.240	22.270	0.000
Covariance	3.647	5.483	0.665	0.506

Chapter 5

Introduction

The purpose of this quantitative study was to determine the language outcomes of retained and matched-promoted English language learners seven years after the initial year of the study. This chapter includes a discussion of results, implications for policy and practice, and limitations to help answer and explore the research question:

> Does grade retention influence language development in English learners, as measured by WIDA Listening, Reading and Writing scale scores, seven years after grade retention?

The study utilized longitudinal data to track retained students, plus those who were post-hoc matched using propensity score matching, in order to create comparable groups. Next, the fixed effects were estimated through use of a multilevel growth model in MPlus. Language development outcomes were negative and significant for the three dependent variables: listening, reading and writing. The null hypothesis was rejected; grade retention negatively influenced all three language development outcomes seven years after retention. English learners who were retained in grade were significantly and demonstrably lower in Listening, Reading and Writing than their continuously promoted counterparts at the end of a seven-year period. This is a brand new finding which contributes to the literature about the outcomes specifically for English language learners.

EL students are already a vulnerable population in US schools, arriving with more obstacles to overcome than non-EL students, such as the stress of learning in a new and unfamiliar language, less access (on average) to social benefits, and added stress if the student's

family has recently immigrated (Crosnoe & Turley, 2011; Yoshikawa & Kalil, 2011; Fix & Zimmerman, 2001). On top of these added difficulties, English learners are more likely to be retained in grade (Tingle, Schoeneberger & Algozzine, 2012), a practice that is strongly linked to dropping out of high school (Hughes, West, Kim & Bauer, 2018; Hughes, Cao, West, Smith & Cerda, 2017). It is, perhaps, no wonder that EL students have lower graduation rates than English-proficient peers (McFarland, Cui & Stark, 2018).

The current study set out to investigate trends in language development between retained and matched-promoted EL students. The findings indicate a clear message: there is no evidence that retention is beneficial to EL students. Grade retention was harmful to student language achievement.

Study Findings, Past and Future Research

The present study investigated the language development outcomes of retained and promoted English learners over seven years, adding to a growing body of knowledge about English learners.

Influence of Retention, Grade and Relative Age

Retention. Results demonstrated that retained and matched-promoted students differ in significant ways. Retained students were significantly lower in their listening, reading and writing language development (intercept), when compared to peers of similar background and initial language levels. No significant differences were found for language growth trajectory (slope).

Language development is complex, requiring both academic and socioemotional factors. The academic component of language is content-area language exposure, described by Cummins (1981) as Cognitive Academic Language Proficiency (CALP), which take five to seven years to

develop. CALP is also commonly described as academic language. Four of the five standards assessed by WIDA cover academic language: the language of science, the language of social studies, the language of language arts, and the language of math (Cook, Boals & Lundberg, 2011). In order to attain full English proficiency, a student must be able to use nuanced and complex grammar and advanced and technical vocabulary in expanded discourse. Exposure to content, and use of academic terms in meaningful communication, are crucial requirements of academic language learning.

Beyond meaningful exposure to, and use of, academic language is a more subtle and nuanced factor: the Affective Filter. First put forth by Stephen Krashen as a theory of second language acquisition, the theory of the affective filter offers that language learning happens best in an environment in which the learner is comfortable and feels safe (Krashen, 1982). When the learner is stressed, the affective filter is an emotional barrier preventing the learning of language.

Numerous studies have explored both academic and socio-emotional outcomes after grade retention. Repeating a grade has been found to be a wash at best for academic achievement (Fruehwirth, Navarro & Takahashi, 2016; Vandecandelaere, Vansteelandt, De Fraine & Van Damme, 2015; Im, Hughes, Kwok, Puckett & Cerda, 2013; Moser, West & Hughes, 2012), with students achieving at roughly the same level as compared to promoted peers upon arrival to middle school. Many other studies have shown that students are worse off academically than their matched peers who were continually promoted (Diris, 2017; Cooley, Navarro & Takahashi, 2011). Thus, grade retention may or may not harm average learners academically upon arrival to middle school. Prior to the present project, one study had indicated that negative academic outcomes of retention might be worse for EL students (Wu, West & Hughes, 2008). This study is

consistent with the prior literature, and greatly expands on the outcomes of retention specifically for English learners.

Whether retention helps or harms the average learner, retained students experience the loss of one year before arrival to middle school. The at-best ineffectiveness of retention alone should be cause for continuous promotion of students: it is an expensive, ineffective and unnecessary course of action even before additional considerations. A number of studies have gone beyond these concerns to explore the socio-emotional outcomes for retained students. As with academic achievement, at best, grade retention is a wash for students emotionally (Cham, Hughes, West & Im, 2015); however, a number of studies have suggested ill effects after retention, including increased misbehavior (Ozek, 2015; Demanet & Van Houtte, 2013), lower confidence (Huddleston & Lowe, 2014; Martin, 2011; Penna & Tallerico, 2005), and school engagement (Im, Hughes, Kwok, Puckett & Cerda, 2013).

Knowing that grade retention can be harmful for both academic and socio-emotional outcomes, and with consideration of Krashen's affective filter theory, it is perhaps no great leap that retention would hinder second language learning. For example, some studies have connected anxiety to development of reading or other second language skills (Bernal Castañeda, 2017; Mohammadpur & Ghafournia, 2015). The present study gives a new dimension of support to Krashen's Affective Filter theory for second language acquisition, as it demonstrates outcomes specifically for retained and matched-promoted English learners.

Initial grade. Students in the study were in kindergarten, first and second grade in 2009, the first year of the study. Students who were initially in kindergarten were significantly lower in language levels for listening, reading and writing. For example, the Listening scale score for a student who was in kindergarten in 2009 were, on average, 15.81 points lower, as compared with

a student who was in first or second grade that year. However, students starting out in kindergarten also indicated significantly higher growth rates in all three domains.

Prior literature is inconsistent on this matter; some studies have found no difference for the age of retention (Silberglitt, Jimerson, Burns & Appleton, 2006), while others have found that later retention (grades 3-6) was worse than early retention in the early grades (Ou & Reynolds, 2010). The present study only compared early grades, so this finding may be unique and certainly warrants consideration of future research.

Relative age. The age-grade comparison variable is derived from district policy indicating that students could enter kindergarten if they were age five on September 1 of the current school year. The target group in the sample reflected students of average age for their grade (5 years old in kindergarten, 6 years in first, 7 years in second) in the first year of the study.

Students who were of average relative age for their grade had higher ending scale scores for Listening, Reading and Writing, as compared with student who were older for grade. For example, a student who was the expected age for his or her grade in 2009 averaged 4.693 Writing scale score points higher than student who were older for their grade. There were also small but significant increases in growth rates for students of average age in Reading and Writing.

English learners displayed higher ending scale scores for listening, reading and writing, plus slightly higher growth in reading and writing, as compared with students who were one or two years older. No students in the sample were young for their grade. This is consistent with current literature, which indicates that students who are older for their cohort may perform lower academically (Martin, 2009).

Influence of Individual Characteristics

Absenteeism. Students who were chronically absent in the 2009-2010 school year were coded as a holdout group for absenteeism. This variable was not significant in most cases, except for a small but significant effect, which indicated that students with a pattern of absenteeism averaged nearly 2 scale score points higher than their counterparts in reading growth.

Gender. In most cases, gender did not indicate significant differences for students in the population. As compared to female counterparts, males achieved similar results for listening and reading for both ending scores and growth trajectory. Significant differences emerged in writing, where males demonstrated statistically less growth, and ended at a lower scale score on average, as compared to females.

Free/ reduced lunch (Socioeconomic Status). Students receiving a free or reduced lunch at school were coded to indicate low socioeconomic status in 2009. Results were not significant in Listening and Writing; however, Reading was significant for intercept, and nearly significant ($p < .1$) for slope. Students receiving a free or reduced lunch in 2009 indicated a slightly higher intercept and growth rate in Reading. This finding is unique, in that the assessment utilized is a measurement of reading comprehension and academic language, so it is difficult to compare to other measures. Some studies have shown that low socioeconomic status is often associated with lower achievement in reading (Hemmerechts, Agirdag & Kavadias, 2017; Lundberg, Larsman & Strid, 2012) and vocabulary skills (Neuman, Kaefer & Pinkham, 2018; Jiang, Logan & Jia, 2018; Morgan, Farkas, Hillemeier, Hammer & Maczuga, 2015), however, newer literature has emerged to suggest great variance in such trends (Sperry, Sperry & Miller, 2018; Gilkerson et al., 2017).

Born outside of the United States. Students identified by a parent/guardian as being born outside of the United States scored slightly but significantly lower in writing for intercept .

Growth rate was negative and almost significant ($p < .1$). This may be supportive of Crosnoe and Lopez-Turley's immigrant paradox (2011). In their work, they describe newly-immigrated students successfully seeking the American Dream, and the paradox that students who start significantly behind their peers tend to display wild success; however, the authors expand upon this idea, specifying that new immigrants achieving the American Dream typically arrive from Africa and Asia; students who are new arrivals from Latin American countries tend to struggle, possibly in response to a multitude of compounding factors, which make life after immigration difficult. School is not an equalizer for these students, who may struggle to catch up for an entire generation (Yoshikawa & Kalil, 2011; De Feyter and Winsler, 2009).

Influence of Baseline Measures

Special Education status. A number of students were on an Individual Education Plan (IEP) in the year before retention; in fact, though rates vary state to state, up to 17% of English language learners are also in Special Education programs nationwide (Counts, Katsiyannis & Whitford, 2018) . At the end of the seven years, dually identified students demonstrated significantly slower growth in listening, reading and writing, plus significantly lower ending scores in Listening, but not Reading or Writing. The indication of a different growth trajectory is consistent with previous research that indicates that dually identified students (English learners on an Individual Education Plan) may experience slower language development as compared with their non-disabled English learner counterparts (Haas et al., 2016).

Listening and Writing Proficiency Level

Students were assessed in the first year of the study for Listening, Speaking, Reading and Writing. Proficiency Level scores were utilized in the propensity score matching process in order to pair students with similar language ability. In the final model, only Listening and Writing

scores were used, as there was interaction between the four domains. Listening was retained to represent both receptive language and verbal skills, while Writing was utilized to represent literacy and expressive language.

Listening and Writing domains were significant for all three outcomes, except Writing slope, which was almost significant ($p < .1$). These results indicate a relationship between starting levels and subsequent language progress. For all three domains, the intercept was positive and significant, while the slope was negative and significant in most cases.

Influence of School Context

School variables. Four school variables were included at the individual level. Teacher experience, school mobility, and percentage of Hispanic students in the school were not significant for any domain; however, attending a school that had met third grade reading benchmarks was significant in two measures. Students in schools that succeeded on third grade measures scored lower in Reading levels (intercept) and Writing growth.

Implications for Policy and Practice

The finding that English learners experienced lower Listening, Reading and Writing levels after grade retention indicates several important applications to practice and policy. Grade retention has been shown to be harmful, or neutral at best, for students academically (Fruehwirth, Navarro & Takahashi, 2016; Vandecandelaere, Vansteelandt, De Fraine & Van Damme, 2015). English learners have a higher likelihood to struggle in school and eventually drop out; this is reason enough to avoid grade retention. Grade retention compounds the challenges that English learners already face in school and is an inappropriate intervention for this population, and, arguably, for all students.

Preventing grade retention for English learners is a systematic change, and cannot happen without planning and education at various levels. Teachers make grade retention decisions based on intuition more often than scholarly findings (Silberglitt, Appleton, Burns & Jimerson, 2006; Bartels, 2003; Tomchin & Impara, 1992); one study indicated that only 9% of teachers making retention decisions attributed their choices to research (Witmer, Hoffman & Nottis, 2004). At the same time, retention decisions are most often made in hopes of improving academic performance (Range, Pijanowski, Holt & Young, 2012; Rafoth & Carey, 1991). Misunderstandings likely arise due to the temporary boost that students experience in their first one to two years after retention (Moser, West & Hughes, 2012; Dong, 2010; Greene & Winters, 2007).

The literature merits a serious consideration of praxis at the school level. First, teacher preparation programs at the university and professional develop at the school level must expand to include a deeper understanding of retention literature, including why grade retention is particularly alarming and likely harmful for English learners. At a minimum, teachers should understand the bigger picture of outcomes after grade retention, including the initial boost that student receive, but more broadly, the outcomes after a number of years, including the damage to language achievement after a number of years.

Schools, local education agencies, and state agencies must also take steps to prevent retention at a higher level. Lacking a research-based retention policy or providing unclear guidelines means subjective decision-making at the teacher level (Beebe-Frankenberger, Bocian, MacMillan & Gresham, 2004). Enacting a policy around grade retention is a crucial step for two reasons: a clear stance against retention as an intervention may prevent retention in the short term, but there is also some indication that a policy stance sends a broader message to teachers and stakeholders (Goos, Schreier, Knipprath, De Fraine, Van Damme & Trautwein, 2013). In

other words, educators' views of the efficacy of the practice are likely susceptible to influence by a local or state policy that does not recognize retention as an effective intervention.

State-level policies with mandatory retention have been called a violation of professional educator standards (Penfield, 2010) and found to be ineffective for improving outcomes (Belot & Vandenberghe, 2014); as new research emerges, this becomes an imperative. Test-based retention more often targets students of color and those whose families have less social capital (LiCalsi, Ozek & Figlio, 2016; Stearns, 2007), including English learners (Tingle, Schoeneberger & Algozzine, 2012; Winsler et al., 2012; Willson & Hughes, 2006). These factors alone, coupled with evidence of a lack of efficacy in improving academics, and a higher dropout rates for English learners, should be enough to abolish such antiquated policies. With the added knowledge that retention is harmful to the English development of language learners, it is clear that test-based retention policies should only exist in history.

Schools and communities can also take steps to prevent retention. For English learners specifically, bilingual education could be helpful decreasing a student's likelihood for being retained (Curiel, Rosenthal & Richek, 1986). Unfortunately, since No Child Left Behind and the emphasis on passing tests in English, funding and support for bilingual programs has dwindled (Lee & Wright, 2014), while pressure to retain "failing" students has increased (Darling-Hammond, 2004). Draining bilingual supports from schools has had a counterproductive effect, as bilingual education is a very strong predictor of student success (Seid, 2016; May, 2014; Thomas and Collier, 2002).

Pre-kindergarten attendance is strongly associated with decreasing or preventing retention (Gormley, Phillips & Anderson, 2018; Dodge, Bai, Ladd & Muschkin, 2017; McCoy et

al., 2017; Winsler et al., 2012), as is attending a program such as Head Start (Phillips, Gormley & Anderson, 2016) or full-day kindergarten (Cannon, Jacknowitz & Painter, 2011).

Limitations

A matching procedure was used to estimate matches between retained and promoted students ahead of comparison. The PSM method reduces selection bias by statistically matching groups post hoc. Background variables were utilized once again in the final model. Such methods are common in human research where ethics prohibit selection of students into a treatment group that has potential negative effects. Though analyses indicate that groups were balanced well, the PSM process is one step below the gold standard of research (random selection). In addition, the matching process required elimination of some cases in both the control and treatment groups. This was necessary in order to create the best matches; however, it caused the treatment group to be substantially lower. It is not optimal to lose cases from the treatment group; future research with larger groups could improve the matching process.

This study also included only students who had complete data for the seven years of the study, meaning that students who tested out during the seven-year period of the study were necessarily excluded. Students were also selected from one urban district in the Midwest, which might mean that there may be limited generalizability for other regions in the United States, and especially in other countries, where the education system is substantially different.

Future Research

The findings in this study open the door to many additional questions relating to the language development of English learners in schools. The study design mirrored modern studies in design, utilizing a propensity score matching approach to compare students seven years after

grade retention. This is referred to in some literature as same-age comparison, which is to say, all students were evaluated seven years after the initial measurement. This design is supported by the fact that the WIDA assessment is vertically aligned, allowing the investigator to compare scale scores across grades. Though early grade retention studies utilized PSM to conduct a same-age comparisons (Hong & Raudenbush, 2005), it is also very common to compare student achievement in the same grade (Hughes, Cao, West, Smith & Cerda, 2017; Steiner, Park & Kim, 2016; Klapproth et al., 2016; Moser, West & Hughes, 2012).

The scope of future study could be expanded in both focus and sample. Investigations might include an examination of Speaking, the domain that was not included in the study, plus math and reading outcomes specifically for English learners after grade retention. A different sample, including other regions of the country, other types of districts, alternative language assessments, and large sample size would also be useful for expanding the body of knowledge, in order to better parse apart the effects that school, and specifically retention, may have for English learners.

Conclusion

Grade retention is a common practice in US schools. Decades of research indicate that grade retention may be either neutral or harmful to students academically and socially, plus may cause students to drop out of high school. English learners have a higher risk for being retained in grade; these factors create myriad disadvantages to a vulnerable population, and indicate an urgency for scholarly work to inform practice.

This study investigated the language development (Listening, Reading and Writing) of matched groups of English learners seven years after grade retention or promotion. Findings

indicated that students who were retained had lower scale scores for Listening, Reading and Writing at the end of the seven-year period. There was no significant difference for slope.

This work indicates that the language development of English learner students is harmed by grade retention. Matched promoted students, whose starting levels were not statistically different, had significantly higher scores in Listening, Reading and Writing at the Year 4 intercept point. Grade retention was not beneficial to EL students; in fact, it was harmful to their language outcomes. Repeating a year of school did not help students' language development.

These findings are important for practitioners. Retention has been shown to be harmful or, at best, neutral for academic and socio-emotional outcomes, and consistently negative for school completion; however, it was unknown whether retention could be helpful to EL students linguistically. The current study contributes to the body of knowledge by demonstrating that language development outcomes are not different from many academic and affective investigations. Grade retention does not benefit English learners' language. This antiquated practice has no benefit to students and cannot replace effective intervention.

References

Abidin, R. R., Golladay, W. M., & Howerton, A. L. (1972). Elementary school retention: An unjustifiable, discriminatory and noxious educational policy. Journal of School Psychology, 9(4), 410-417.

Akmal, T. T., & Larsen, D. E. (2004). Keeping history from repeating itself: Involving parents about retention decisions to support student achievement. RMLE Online, 27(2), 1-14.

Allen, C. S., Chen, Q., Willson, V. L., & Hughes, J. N. (2009). Quality of Research Design Moderates Effects of Grade Retention on Achievement: A Meta-Analytic, Multilevel Analysis. Educational Evaluation And Policy Analysis, 31(4), 480-499.

Allensworth, E. M. (2005). Dropout Rates after High-Stakes Testing in Elementary School: A Study of the Contradictory Effects of Chicago's Efforts to End Social Promotion. Educational Evaluation And Policy Analysis, 27(4), 341-364.

Anfinson, R. D. (1941). School progress and pupil adjustment. The Elementary School Journal, 41(7), 507-514.

Antecol, H., Eren, O., & Ozbeklik, S. (2016). Peer Effects in Disadvantaged Primary Schools Evidence from a Randomized Experiment. Journal of Human Resources, 51(1), 95-132.

Austin, P. C. (2011a). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behavioral Research, 46(3), 399-424.

Austin, P. C. (2011b). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. Pharmaceutical statistics, 10(2), 150-161.

Ayres, L. P. (1909). Laggards in our schools.

Bacallao, M. L., & Smokowski, P. R. (2007). The Costs of Getting Ahead: Mexican Family System Changes after Immigration. Family Relations, 56(1), 52-66.

Bartels, N. (2003). How teachers and researchers read academic articles. Teaching and Teacher Education, 19(7), 737-753.

Beebe-Frankenberger, M., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Sorting Second-Grade Students: Differentiating Those Retained From Those promoted. Journal Of Educational Psychology, 96(2), 204-215.

Belot, M., & Vandenberghe, V. (2014). Evaluating the "Threat" Effects of Grade Repetition: Exploiting the 2001 Reform by the French-Speaking Community of Belgium. Education Economics, 22(1), 73-89.

Bernal Castañeda, S. (2017). Affective Limitations in Second Language Acquisition by Spanish Adult Learners in Vocational Training Programs. Latin American Journal of Content and Language Integrated Learning, 10(1), 133–160.

Bowman, L. J. (2005). Grade Retention: Is It a Help or Hindrance to Student Academic Success?. Preventing School Failure, 49(3), 42-46.

Braun, D. L., Gable, R. K., Billups, F. D., Vieira, M., & Blasczak, D. (2016). Evolving an Accelerated School Model through Student Perceptions and Student Outcome Data. Journal of Educational Issues, 2(2), 385-406.

Brown v. Board of Education, 347 U.S. 483 (1954).

Burkam, D. T., LoGerfo, L., Ready, D., & Lee, V. E. (2007). The differential effects of repeating kindergarten. Journal of Education for Students Placed at Risk, 12(2), 103-136.

Burke, M. A., & Sass, T. R. (2013). Classroom peer effects and student achievement. Journal of Labor Economics, 31(1), 51-82.

Bush, G. W. (2001). Address to a joint session of Congress and the American people. Harv. JL & Pub. Pol'y, 25, xviii.

Byrd, R. S., & Weitzman, M. L. (1994). Predictors of Early Grade Retention among Children in the United States. Pediatrics, 93(3), 481-87.

Cannon, J. S., Jacknowitz, A., & Painter, G. (2011). The Effect of Attending Full-Day Kindergarten on English Learner Students. Journal Of Policy Analysis And Management, 30(2), 287-309.

Carlton, M. P., & Winsler, A. (1999). School readiness: The need for a paradigm shift. School Psychology Review, 28(3), 338.

Castañeda v. Pickard, 648 F.2d 989 (5th Cir. 1981).

Cham, H., Hughes, J. N., West, S. G., & Im, M. H. (2015). Effect of retention in elementary grades on grade 9 motivation for educational attainment. Journal of school psychology, 53(1), 7-24.

Chen, Q., Hughes, J. N., & Kwok, O. (2014). Differential Growth Trajectories for Achievement among Children Retained in First Grade: A Growth Mixture Model. Elementary School Journal, 114(3), 327-353.

Chen, X., Liu, C., Zhang, L., Shi, Y., & Rozelle, S. (2010). Does Taking One Step Back Get You Two Steps Forward? Grade Retention and School Performance in Poor Areas in Rural China. International Journal Of Educational Development, 30(6), 544-559.

Cisneros v. Corpus Christi Independent School District, 330 F. Supp. 1377 S.D. Tex. (1971)

Clinton, W. J. (1998). President William Jefferson Clinton State of the Union Address. The White House, Office of the Press Secretary, 19.

Chicago

Clinton, W. J. (1999). President William Jefferson Clinton State of the Union Address. The White House, Office of the Press Secretary, 19.

Chicago

Cochran, W. G., & Rubin, D. B. (1973). Controlling bias in observational studies: A review. Sankhyā: The Indian Journal of Statistics, Series A, 417-446.

Coffield, W. H., & Blommers, P. (1956). Effects of non-promotion on educational achievement in the elementary school. Journal of Educational Psychology, 47(4), 235.

Contreras, A. R., & Valverde, L. A. (1994). The impact of Brown on the education of Latinos. Journal of Negro Education, 470-481.

Cook, H. G., Boals, T., & Lundberg, T. (2011). Academic Achievement for English Learners: What Can We Reasonably Expect?. Phi Delta Kappan, 93(3), 66-69.

Cook, H. G., Boals, T., & Lundberg, T. (2011). What can we reasonably expect?. Phi Delta Kappan, 93(3), 66.

Cook, H., Wilmes, C., Boals, T., Santos, M., & Wisconsin Center for Education Research, M. n. (2008). Issues in the Development of Annual Measurable Achievement Objectives for WIDA Consortium States. WCER Working Paper No. 2008-2. Wisconsin Center For Education Research.

Cook, W. W. (1941). Some effects of the maintenance of high standards of promotion. The Elementary School Journal, 41(6), 430-437.

Cooley, J., Navarro, S., & Takahashi, Y. (2011). How the timing of grade retention affects outcomes: Identification and estimation of time-varying treatment effects.

Counts, J., Katsiyannis, A., & Whitford, D. K. (2018). Culturally and Linguistically Diverse Learners in Special Education: English Learners. NASSP Bulletin, 102(1), 5-21.

Crosnoe, R., & Turley, R. (2011). K-12 Educational Outcomes of Immigrant Youth. Future Of Children, 21(1), 129-152.

Cummins, J. (1981). Empirical and theoretical underpinnings of bilingual education. Journal of education, 16-29.

Curiel, H., Rosenthal, J. A., & Richek, H. G. (1986). Impacts of bilingual education on secondary school grades, attendance, retentions and drop-out. Hispanic Journal Of Behavioral Sciences, 8(4), 357-367.

Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of 'No Child Left Behind'. Race Ethnicity and Education, 10(3), 245-260.

Darling-Hammond, L. (2006). No Child Left Behind and High School Reform. Harvard Educational Review, 76(4), 642-667.

Darling-Hammond, L. (2004). Standards, Accountability, and School Reform. Teachers College Record, 106(6), 1047-1085.

De Feyter, J. J., & Winsler, A. (2009). The early developmental competencies and school readiness of low-income, immigrant children: Influences of generation, race/ethnicity, and national origins. Early Childhood Research Quarterly, 24(4), 411-431.

De Witte, K., Nicaise, I., Lavrijsen, J., Van Landeghem, G., Lamote, C., & Van Damme, J. (2013). The Impact of Institutional Context, Education and Labour Market Policies on Early School Leaving: A Comparative Analysis of EU Countries. European Journal Of Education, 48(3), 331-345.

Del Rio Independent School District v. Salvatierra, 33 S.W.2d 790 (1930)

Delgado v. Bastrop Independent School District, 388 (1948).

Demanet, J., & Van Houtte, M. (2013). Grade retention and its association with school misconduct in adolescence: A multilevel approach. School Effectiveness and School Improvement, 24(4), 417-434.

Demanet, J., & Van Houtte, M. (2016). Are Flunkers Social Outcasts? A Multilevel Study of Grade Retention Effects on Same-Grade Friendships. American Educational Research Journal, 53(3), 745-780.

Department of Education (2008). A Nation Accountable: Twenty-Five Years after "A Nation at Risk". US Department Of Education.

Dillon, S. (2010). Foreign languages fade in class–except Chinese. The New York Times, A18.

Diris, R. (2017). Don't Hold Back? The Effect of Grade Retention on Student Achievement. Education Finance and Policy, 12(3), 312-341.

Dobbs, V., & Neville, D. (1967). The Effect of Nonpromotion on the Achievement of Groups Matched from Retained First Graders and Promoted Second Graders. The Journal of Educational Research, 60(10), 472-475.

Dodge, K. A., Bai, Y., Ladd, H. F., & Muschkin, C. G. (2017). Impact of North Carolina's Early Childhood Programs and Policies on Educational Outcomes in Elementary School. Child development, 88(3), 996-1014.

Dodge, K. A., Bai, Y., Ladd, H. F., & Muschkin, C. G. (2017). Impact of North Carolina’s Early Childhood Programs and Policies on Educational Outcomes in Elementary School. Child Development, 88(3), 996–1014.

Dombek, J., & Connor, C. (2012). Preventing Retention: First Grade Classroom Instruction and Student Characteristics. Psychology In The Schools, 49(6), 568-588.

Dong, Y. (2010). Kept back to get ahead? Kindergarten retention and academic performance. European Economic Review, 54(2), 219-236.

Dougan, K., & Pijanowski, J. (2011). The Effects of Academic Redshirting and Relative Age on Student Achievement. International Journal Of Educational Leadership Preparation, 6(2),

Ehmke, T., Drechsel, B., & Carstensen, C. H. (2010). Effects of grade retention on achievement and self-concept in science and mathematics. Studies in Educational Evaluation, 36(1), 27-35.

El-Hassan, K. (1998). Relation of Academic History and Demographic Variables to Grade Retention in Lebanon. Journal Of Educational Research, 91(5), 279-88.

Every Student Succeeds Act of 2015, Pub. L. No. 114-95 § 114 Stat. 1177 (2015-2016).

Feld, J., & Zölitz, U. (2017). Understanding peer effects: on the nature, estimation, and channels of peer effects. Journal of Labor Economics, 35(2), 387-428.

Ferguson, P. C. (1991). Longitudinal Outcome Differences among Promoted and Transitional At-Risk Kindergarten Students. Psychology In The Schools, 28(2), 139-46.

Fix, M., & Zimmerman, W. (2001). All under One Roof: Mixed-Status Families in an Era of Reform. International Migration Review, 35(2), 397-419.

Fortner, C. K., & Jenkins, J. M. (2016). Kindergarten redshirting: Motivations and spillovers using census-level data. Early Childhood Research Quarterly, 38, 44-56.

Fruehwirth, J. C., Navarro, S., & Takahashi, Y. (2016). How the timing of grade retention affects outcomes: Identification and estimation of time-varying treatment effects. Journal of Labor Economics, 34(4), 979-1021.

García, O. (2014). US Spanish and Education Global and Local Intersections. Review of Research in Education, 38(1), 58-80.

Gelatt, J. & Zong, J. (2018). Settling in: A profile of the unauthorized immigrant population in the United States. Migration Policy Institute. Retrieved from https://www.migrationpolicy.org/research/profile-unauthorized-immigrant-population-united-states.

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Kimbrough Oller, D. & Paul, T. D. (2017). Mapping the early language environment using all-day recordings and automated analysis. American Journal of Speech-Language Pathology, 26(2), 248-265.

Gilliam, W. S., & Zigler, E. F. (2000). A Critical Meta-Analysis of All Evaluations of State-Funded Preschool from 1977 to 1998: Implications for Policy, Service Delivery and Program Evaluation. Early Childhood Research Quarterly, 15(4), 441-73.

Goldstein, J., Eastwood, M., & Behuniak, P. (2014). Can Teacher Ratings of Students' Skills at Kindergarten Entry Predict Kindergarten Retention?. Journal Of Educational Research, 107(3), 217-229.

Goos, M., Schreier, B., Knipprath, H., De Fraine, B., Van Damme, J., & Trautwein, U. (2013). How Can Cross-Country Differences in the Practice of Grade Retention Be Explained? A Closer Look at National Educational Policy Factors. Comparative Education Review, 57(1), 54-84.

Goos, M., Van Damme, J., Onghena, P., Petry, K., & de Bilde, J. (2013). First-grade retention in the Flemish educational context: Effects on children's academic growth, psychosocial growth, and school career throughout primary education. Journal of school psychology, 51(3), 323-347.

Gormley Jr, W. T., Phillips, D., & Anderson, S. (2018). The Effects of Tulsa's Pre-K Program on Middle School Student Performance. Journal of Policy Analysis and Management, 37(1), 63-87.

Greene, J. P., & Winters, M. A. (2007). Revisiting Grade Retention: An Evaluation of Florida's Test-Based Promotion Policy. Education Finance And Policy, 2(4), 319-340.

Greene, J. P., & Winters, M. A. (2009). The Effects of Exemptions to Florida's Test-Based Promotion Policy: Who Is Retained?--Who Benefits Academically?. Economics Of Education Review, 28(1), 135-142.

Greenfield, G. A., & Kates Jr, D. B. (1975). Mexican Americans, Racial Discrimination, and the Civil Rights Act of 1866. California Law Review, 662-731.

Haas, E., Huang, M., Tran, L., Yu, A., Regional Educational Laboratory West (ED), National Center for Education Evaluation and Regional Assistance (ED), & WestEd. (2016). The Achievement Progress of English Learner Students in Utah. REL 2016-155. Regional Educational Laboratory West. Regional Educational Laboratory West.

Hansen, B. B., & Bowers, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. Statistical Science, 219-236.

Hart, B., & Risley, T. R. (1992). American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. Developmental psychology, 28(6), 1096.

Heck, R. H., & Thomas, S. L. (2015). An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus. Routledge.

Heck, R. H., Tabata, L., & Thomas, S. L. (2013). Multilevel and longitudinal modeling with IBM SPSS. Routledge.

Hemmerechts, K., Agirdag, O., & Kavadias, D. (2017). The Relationship between Parental Literacy Involvement, Socio-Economic Status and Reading Literacy. Educational Review, 69(1), 85–101.

Holland, P. W. (1986). Statistics and causal inference. Journal of the American statistical Association, 81(396), 945-960.

Holmes, C. T., & Matthews, K. M. (1984). The Effects of Nonpromotion on Elementary and Junior High School Pupils: A Meta-Analysis. Review Of Educational Research, 54(2), 225-36.

Hong, G., & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. Educational Evaluation and Policy Analysis, 27(3), 205-224.

Hong, G., & Yu, B. (2007). Early-grade retention and children's reading and math learning in elementary years. Educational Evaluation and Policy Analysis, 29(4), 239-261.

Huddleston, A. P. (2015). "Making the Difficult Choice": Understanding Georgia's Test-Based Grade Retention Policy in Reading. Education Policy Analysis Archives, 23(51),

Huddleston, A. P., & Lowe, T. N. (2014). "I Skim and Find the Answers": Addressing Search-and-Destroy in Reading. Reading Teacher, 68(1), 71-79.

Hughes, J. N., Cao, Q., West, S. G., Smith, P. A., & Cerda, C. (2017). Effect of retention in elementary grades on dropping out of school early. Journal of School Psychology, 65, 11-27.

Hughes, J. N., Chen, Q., Thoemmes, F., & Kwok, O. (2010). An Investigation of the Relationship between Retention in First Grade and Performance on High Stakes Tests in Third Grade. Educational Evaluation And Policy Analysis, 32(2), 166-182.

Hughes, J. N., West, S. G., Kim, H., & Bauer, S. S. (2018). Effect of Early Grade Retention on School Completion: A Prospective Study. Journal of Educational Psychology, 110(7), 974–991.

Hunt, J. W. (2008). "A Nation at Risk" and No Child Left Behind: Deja Vu for Administrators?. Phi Delta Kappan, 89(8), 580-585.

Im, M. H., Hughes, J. N., Kwok, O. M., Puckett, S., & Cerda, C. A. (2013). Effect of retention in elementary grades on transition to middle school. Journal of school psychology, 51(3), 349-365.

Hughes, J. N., Chen, Q., Thoemmes, F., & Kwok, O. (2010). An Investigation of the Relationship between

Jacob, B. A., & Lefgren, L. (2009). The effect of grade retention on high school completion. American Economic Journal: Applied Economics, 33-58.

Jiang, H., Logan, J. A., & Jia, R. (2018). Modeling the Nature of Grammar and Vocabulary Trajectories from Prekindergarten to Third Grade. Journal of Speech, Language, and Hearing Research, 61(4).

Jimerson, S. D. (2002). Winning the battle and losing the war: Examining the realation between grade retention and dropping out of high school. Psychology In The Schools, 39(4), 441.

Jimerson, S. R. (1999). On the failure of failure: Examining the association between early grade retention and education and employment outcomes during late adolescence. Journal of School Psychology, 37(3), 243-272.

Jimerson, S. R., & Ferguson, P. (2007). A Longitudinal Study of Grade Retention: Academic and Behavioral Outcomes of Retained Students through Adolescence. School Psychology Quarterly, 22(3), 314-339.

Jimerson, S. R., Pletcher, S. W., Graydon, K., Schnurr, B. L., Nickerson, A. B., & Kundert, D. K. (2006). Beyond Grade Retention and Social Promotion: Promoting the Social and Academic Competence of Students. Psychology In The Schools, 43(1), 85-97.

Jimerson, S., Carlson, E., Rotert, M., Egeland, B., & Sroufe, L. A. (1997). A prospective, longitudinal study of the correlates and consequences of early grade retention. Journal of School Psychology, 35(1), 3-25.

Karoly, L., & Gonzalez, G. (2011). Early Care and Education for Children in Immigrant Families. Future Of Children, 21(1), 71-101.

Kenyon, D. M., MacGregor, D., Li, D., & Cook, H. G. (2011). Issues in vertical scaling of a K-12 English language proficiency test. Language Testing, 28(3), 383-400.

Keyes, C. H. (1911). Progress Through the Grades of City Schools: A Study of Acceleration and Arrest (No. 42). Teachers College, Columbia University.

Kim, S., Mazza, J., Zwanziger, J., & Henry, D. (2014). School and Behavioral Outcomes among Inner City Children: Five-Year Follow-Up. Urban Education, 49(7), 835-856.

Klapproth, F., Schaltz, P., Brunner, M., Keller, U., Fischbach, A., Ugen, S., & Martin, R. (2016). Short-term and medium-term effects of grade retention in secondary school on academic achievement and psychosocial outcome variables. Learning and Individual Differences, 50, 182-194.

Krashen, S. (2015) The ELL" Problem" and the Common Core" Solution". Retrieved from http://skrashen.blogspot.com/2015/07/the-ell-problem-and-common-core-solution.html

Krashen, S. D. (1987). Principles and practice in second language acquisition (pp. 1982-1982). New York.

Lamote, C., Pinxten, M., Van Den Noortgate, W., & Van Damme, J. (2014). Is the Cure Worse than the Disease? A Longitudinal Study on the Effect of Grade Retention in Secondary Education on Achievement and Academic Self-Concept. Educational Studies, 40(5), 496-514.

Lau v. Nichols, 414 U.S. 563, 94 S. Ct. 786, 39 L. Ed. 2d 1 (1974).

Lavrijsen, J., & Nicaise, I. (2017). Systemic Obstacles to Lifelong Learning: The Influence of the Educational System Design on Learning Attitudes. Studies in Continuing Education, 39(2), 176–196.

Lazarus, P. J., & Ortega, P. (2007). Universal Pre-Kindergarten in Conjunction with Universal Screenings: An Antidote to Grade Retention. Journal Of Educational Research & Policy Studies, 7(1), 54-75.

Lee, J. S., & Wright, W. E. (2014). The Rediscovery of Heritage and Community Language Education in the United States. Review of Research in Education, 38(1), 137-165.

Lesaux, N. K. (2012). Reading and reading instruction for children from low-income and non-English-speaking households. The Future of Children, 22(2), 73-88.

LiCalsi, C., Ozek, U., & Figlio, D. (2016). The Uneven Implementation of Universal School Policies: Maternal Education and Florida's Mandatory Grade Retention Policy. Society for Research on Educational Effectiveness.

Lin, G. C. (2008). Pedagogies Proving Krashen's Theory of Affective Filter. Online Submission.

Lindvig, E. K. (1983). Grade Retention: Evolving Expectations and Individual Differences. Clearing House, 56(6), 253-56.

London, R. A., Sanchez, M., & Castrechini, S. (2016). The dynamics of chronic absence and student achievement. education policy analysis archives, 24, 112.

Lundberg, I., Larsman, P., & Strid, A. (2012). Development of Phonological Awareness during the Preschool Year: The Influence of Gender and Socio-Economic Status. Reading and Writing: An Interdisciplinary Journal, 25(2), 305–320.

Malone, L. M., West, J., Denton, K. F., Park, J., National Center for Education Statistics (ED), W. D., & Education Statistics Services Inst., W. D. (2006). The Early Reading and Mathematics Achievement of Children Who Repeated Kindergarten or Who Began School a Year Late. Statistics In Brief. NCES 2006-064.

Mann, D. (1986). Can we help dropouts: Thinking about the undoable. The Teachers College Record, 87(3), 307-323.

Mariano, L. T., & Martorell, P. (2013). The Academic Effects of Summer Instruction and Retention in New York City. Educational Evaluation And Policy Analysis, 35(1), 96-117.

Marsh, H. W. (2016). Cross-cultural generalizability of year in school effects: Negative effects of acceleration and positive effects of retention on academic self-concept. Journal of Educational Psychology, 108(2), 256.

Marsh, H. W. (2016). Cross-cultural generalizability of year in school effects: Negative effects of acceleration and positive effects of retention on academic self-concept. Journal of Educational Psychology, 108(2), 256.

Martin, A. J. (2009). Age Appropriateness and Motivation, Engagement, and Performance in High School: Effects of Age within Cohort, Grade Retention, and Delayed School Entry. Journal Of Educational Psychology, 101(1), 101-114.

Martin, A. J. (2011). Holding Back and Holding behind: Grade Retention and Students' Non-Academic and Academic Outcomes. British Educational Research Journal, 37(5), 739-763.

Mathys, C., Véronneau, M. H., & Lecocq, A. (2017). Grade Retention at the Transition to Secondary School: Using Propensity Score Matching to Identify Consequences on Psychosocial Adjustment. The Journal of Early Adolescence, 0272431617735651.

May, D. C., & Welch, E. L. (1984). The effects of developmental placement and early retention on children's later scores on standardized tests. Psychology In The Schools, 21(3), 381-385.

May, S. (2014). Justifying educational language rights. Review of Research in Education, 38(1), 215-241.

McCoy, A. R., & Reynolds, A. J. (1999). Grade Retention and School Performance:: An Extended Investigation. Journal of School Psychology, 37(3), 273-298.

McCoy, D. C., Yoshikawa, H., Ziol-Guest, K. M., Duncan, G. J., Schindler, H. S., Magnuson, K., Shonkoff, J. P. (2017). Impacts of Early Childhood Education on Medium- and Long-Term Educational Outcomes. Educational Researcher, 46(8), 474–487.

McFarland, J., Cui, J. & Stark, P. (2018). Trends in High School Dropout and Completion Rates in the United States: 2014. Compendium Report. NCES 2017-18. National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubs2018/2018117.pdf

McGill-Franzen, A., Zmach, C., Solic, K., & Zeig, J. L. (2006). The Confluence of Two Policy Mandates: Core Reading Programs and Third-Grade Retention in Florida. Elementary School Journal, 107(1), 67-93.

McLeskey, J., Lancaster, M., & Gizzle, K. L. (1995). Learning disabilities and grade retention: A review of issues with recommendations for practice. Learning Disabilities Research & Practice, 10(2), 120-128.

Mendez v. Westminster, 161 F.2d 774 (1947)

Menken, K. (2010). NCLB and English Language Learners: Challenges and Consequences. Theory Into Practice, 49(2), 121-128.

Meyer, E. (2014). The culture map: Breaking through the invisible boundaries of global business. PublicAffairs.

Ming, K., & Rosenbaum, P. R. (2000). Substantial gains in bias reduction from matching with a variable number of controls. Biometrics, 56(1), 118-124.

Mitchell, C., Robertson. K., & Trez, C. (2016, October 6). English-language-learner literacy under the Every Student Succeeds Act [Webinar]. Retrieved from https://webinars.on24.com/edweek/ELL

Mohammadpur, B., & Ghafournia, N. (2015). An Elaboration on the Effect of Reading Anxiety on Reading Achievement. English Language Teaching, 8(7), 206-215.

Morgan, P. L., Farkas, G., Hillemeier, M. M., Hammer, C. S., & Maczuga, S. (2015). 24-Month-Old Children with Larger Oral Vocabularies Display Greater Academic and Behavioral Functioning at Kindergarten Entry. Child Development, 86(5), 1351–1370.

Moser, S. E., West, S. G., & Hughes, J. N. (2012). Trajectories of Math and Reading Achievement in Low-Achieving Children in Elementary School: Effects of Early and Later Retention in Grade. Journal Of Educational Psychology, 104(3), 603-621.

National Center for Education Statistics. (2019). English language learners in public schools. Retrieved from https://nces.ed.gov/programs/coe/indicator_cgf.asp.

National Center for Education Statistics, 2016. English language learners. Retrieved from http://nces.ed.gov/programs/coe/indicator_cgf.asp

Neuman, S. B., Kaefer, T., & Pinkham, A. M. (2018). A Double Dose of Disadvantage: Language Experiences for Low-Income Children in Home and School. Journal of Educational Psychology, 110(1), 102–118.

Office of Educational Quality & Accountability (2010). Oklahoma School Profiles. Retrieved from https://schoolreportcard.org/report-card/district/list

Oklahoma City Public Schools (2014). Statistical Profile 2012-2013. Retrieved from https://www.okcps.org/Domain/115.

Oklahoma City Public Schools (2017). Statistical Profile 2015-2016. Retrieved from https://www.okcps.org/Domain/115.

Oklahoma City Public Schools Board of Education (2005). Student enrollment and residency: Requirements and guardianship, homelessness, and foster care. (OKCPS F-02). Retrieved from https://www.okcps.org/Page/255.

Ou, S., & Reynolds, A. J. (2010). Grade Retention, Postsecondary Education, and Public Aid Receipt. Educational Evaluation And Policy Analysis, 32(1), 118-139.

Özek, U. (2015). Hold back to move forward? Early grade retention and student misbehavior. Education Finance and Policy.

Palmer, J (2018). Oklahoma Nearly Tops Nation in Holding Back Early-Grade Students. Oklahoma Watch. Retrieved from http://oklahomawatch.org/2018/12/14/oklahoma-nearly-tops-nation-in-holding-back-early-grade-students/

Penfield, R. D. (2010). Test-Based Grade Retention: Does It Stand Up to Professional Standards for Fair and Appropriate Test Use?. Educational Researcher, 39(2), 110-119.

Penna, A. A., & Tallerico, M. (2005). Grade Retention and School Completion: Through Students' Eyes. Journal Of At-Risk Issues, 11(1), 13-17.

Peterson, L. S., & Hughes, J. N. (2011). The Differences between Retained and Promoted Children in Educational Services Received. Psychology In The Schools, 48(2), 156-165.

Phillips, D., Gormley, W., & Anderson, S. (2016). The effects of Tulsa's CAP Head Start program on middle-school academic outcomes and progress. Developmental psychology, 52(8), 1247.

Plessy v. Ferguson, 163 U.S. 537 (1896).

Plyler v. Doe, 457 U.S. 202 (1982)

Powers, J. M. (2014). From Segregation to School Finance The Legal Context for Language Rights in the United States. Review of Research in Education, 38(1), 81-105.

Raffaele Mendez, L. M., Kim, E. S., Ferron, J., & Woods, B. (2015). Altering School Progression through Delayed Entry or Kindergarten Retention: Propensity Score Analysis of Long-Term Outcomes. Journal Of Educational Research, 108(3), 186-203.

Rafoth, M. A., & Carey, K. T. (1991). A survey of state-level contacts for school psychology regarding retention/promotion practices: Are we evaluating the risks and benefits?. Psychology In The Schools, 28(1), 35-42.

Ramey, C. T., & Ramey, S. L. (2004). Early Learning and School Readiness: Can Early Intervention Make a Difference?. Merrill-Palmer Quarterly: Journal Of Developmental Psychology, 50(4), 471-491.

Range, B. G., Pijanowski, J., Holt, C. R., & Young, S. (2012). The Perceptions of Primary Grade Teachers and Elementary Principals about the Effectiveness of Grade-Level Retention. Professional Educator, 36(1).

Raphael, S., & Smolensky, E. (2009). Immigration and poverty in the United States. The American Economic Review, 99(2), 41-44.

Reschly, A. L., & Christenson, S. L. (2013). Grade retention: Historical perspectives and new research. Journal of school psychology, 51(3), 319-322.

Resnick, M. D., Bearman, P. S., Blum, R. W., Bauman, K. E., Harris, K. M., Jones, J., ... & Udry, J. R. (1997). Protecting adolescents from harm: findings from the National Longitudinal Study on Adolescent Health. Jama, 278(10), 823-832.

Reynolds, A. J. (1992). Grade Retention and School Adjustment: An Explanatory Analysis. Educational Evaluation And Policy Analysis, 14(2), 101-21.

Rhodes, J., Thomas, J. M., & Liles, A. (2018). Predictors of grade retention among children in an elementary school truancy intervention. Journal of At-Risk Issues, 21(1).

Roderick, M. (1994). Grade Retention and School Dropout: Investigating the Association. American Educational Research Journal, 31(4), 729-59.

Roderick, M., & Nagaoka, J. (2005). Retention Under Chicago's High-Stakes Testing Program: Helpful, Harmful, or Harmless?. Educational Evaluation And Policy Analysis, 27(4), 309-340.

Rodney, L. W., Crafter, B., Rodney, H. E., & Mupier, R. M. (1999). Variables Contributing to Grade Retention among African American Adolescent Males. Journal Of Educational Research, 92(3), 185-90.

Rosenbaum, P. R., & Rubin, D. B. (1985a). The bias due to incomplete matching. Biometrics, 103-116.

Rosenbaum, P. R., & Rubin, D. B. (1985b). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. The American Statistician, 39(1), 33-38.

Rosenbaum, P. R., & Rubin, D. B.. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika, 70(1), 41–55.

Rubin, D. B. (1973a). The use of matched sampling and regression adjustment to remove bias in observational studies. Biometrics, 185-203.

Rubin, D. B. (1973b). Matching to remove bias in observational studies. Biometrics, 159-183.

Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. Statistics in medicine, 26(1), 20-36.

Rubin, D. B., & Thomas, N. (1996). Matching using estimated propensity scores: relating theory to practice. Biometrics, 249-264.

Rueda, R., & Windmueller, M. P. (2006). English Language Learners, LD, and Overrepresentation: A Multiple-Level Analysis. Journal Of Learning Disabilities, 39(2), 99-107.

Rust, J. O., & Wallace, K. A. (1993). Effects of grade level retention for four years. Journal of Instructional Psychology, 20(2), 162.

Schaller, A., Rocha, L., & Barshinger, D. (2007). Maternal Attitudes and Parent Education: How Immigrant Mothers Support Their Child's Education Despite Their Own Low Levels of Education. Early Childhood Education Journal, 34(5), 351-356. doi:10.1007/s10643-006-0143-6

Schwager, M. T., Mitchell, D. E., Mitchell, T. K., & Hecht, J. B. (1992). How school district policy influences grade level retention in elementary schools. Educational Evaluation and Policy Analysis, 14(4), 421-438.

Seid, Y. (2016). Does learning in mother tongue matter? Evidence from a natural experiment in Ethiopia. Economics of Education Review, 55, 21-38.

Sheng, Z. J. (2011). Dropping out of School among ELL Students: Implications to Schools and Teacher Education. Clearing House, 84(3), 98.

Shepard, L. A., & Smith, M. L. (1990). Synthesis of Research on Grade Retention. Educational Leadership, 47(8), 84-88.

Silberglitt, B., Appleton, J. J., Burns, M. K., & Jimerson, S. R. (2006). Examining the effects of grade retention on student reading performance: A longitudinal study. Journal of School Psychology, 44(4), 255-270.

Silberglitt, B., Jimerson, S. R., Burns, M. K., & Appleton, J. J. (2006). Does the Timing of Grade Retention Make a Difference? Examining the Effects of Early versus Later Retention. School Psychology Review, 35(1), 134-141.

Sox, A. K. (2009). Latino Immigrant Students in Southern Schools: What We Know and Still Need to Learn. Theory Into Practice, 48(4), 312-318.

Sperry, D. E., Sperry, L. L., & Miller, P. J. (2018). Reexamining the verbal environments of children from different socioeconomic backgrounds. Child development.

Stearns, E. (2007). Staying Back and Dropping Out: The Relationship Between Grade Retention and School Dropout. Sociology Of Education, 80(3), 210-240.

Steiner, P. M., Park, S., & Kim, Y. (2016). Identifying causal estimands for time-varying treatments measured with time-varying (age or grade-based) instruments. Multivariate behavioral research, 51(6), 865-8780.

Stone, S., & Engel, M. (2007). Same Old, Same Old? Students' Experiences of Grade Retention under Chicago's Ending Social Promotion Policy. American Journal Of Education, 113(4), 605-634.

Stringer, L. A. (1960). Report on a retentions program. The Elementary School Journal, 60(7), 370-375.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. Statistical science: a review journal of the Institute of Mathematical Statistics, 25(1), 1.

TESOL (2015). TESOL International Association Releases Statement on Every Student Succeeds Act [Press Release]. Retrieved from http://www.tesol.org/docs/default-source/press-releases/tesol-statement-on-essa-4-dec-15.pdf?sfvrsn=2.

Thoemmes, F. (2012). Propensity score matching in SPSS. arXiv preprint arXiv:1201.6385.

Thomas, W. P., Collier, V. P., & Center for Research on Education, D. C. (2002). A National Study of School Effectiveness for Language Minority Students' Long-Term Academic Achievement.

Tillman, K. H., Guo, G., & Harris, K. M. (2006). Grade retention among immigrant children. Social Science Research, 35(1), 129-156.

Tingle, L. R., Schoeneberger, J., & Algozzine, B. (2012). Does Grade Retention Make a Difference?. Clearing House: A Journal Of Educational Strategies, Issues And Ideas, 85(5), 179-185.

Tomchin, E. M., & Impara, J. C. (1992). Unraveling teachers' beliefs about grade retention. American Educational Research Journal, 29(1), 199-223.

Tulsa Public Schools, 2017. About TPS. Retrieved from http://www.tulsaschools.org/4_About_District/aboutTPS_main.asp.

Tumlinson, S. E., Sass, D. A., & Cano, S. M. (2014). The search for causal inferences: Using propensity scores post hoc to reduce estimation error with nonexperimental research. Journal of pediatric psychology, 39(2), 246-257.

United States Department of Education National Center for Education Statistics (2014). Dropout rates in the United States. Washington, DC: National Center for Education Statistics.

Van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. Educational Research Review, 5(2), 134-150.

Vandecandelaere, M., Vansteelandt, S., De Fraine, B., & Van Damme, J. (2015). The effects of early grade retention: Effect modification by prior achievement and age. Journal of school psychology, 54, 77-93.

Vandecandelaere, M., Vansteelandt, S., De Fraine, B., & Van Damme, J. (2016). Time-varying treatments in observational studies: Marginal structural models of the effects of early grade retention on math achievement. Multivariate behavioral research, 1-22.

Varela, E. (2010). Mainstreaming ELLs into Grade-Level Classes. Education Digest: Essential Readings Condensed For Quick Review, 76(2), 39-43.

Wang, Y. L., & Johnstone, W. (1997). Evaluation of a Pre-First Transitional Program. ERS Spectrum, 15(3), 40-47.

Warren, J. R., Hoffman, E., & Andrew, M. (2014). Patterns and Trends in Grade Retention Rates in the United States, 1995-2010. Educational Researcher, 43(9), 433-443.

Warren, J., & Saliba, J. (2012). First- through Eighth-Grade Retention Rates for All 50 States: A New Method and Initial Results. Educational Researcher, 41(8), 320-329.

West, S. G., Cham, H., Thoemmes, F., Renneberg, B., Schulze, J., & Weiler, M. (2014). Propensity scores as a basis for equating groups: Basic principles and application in clinical treatment outcome research. Journal of consulting and clinical psychology, 82(5), 906.

Willson, V. L., & Hughes, J. N. (2006). Retention of Hispanic/Latino students in first grade: Child, parent, teacher, school, and peer predictors. Journal of School Psychology, 44(1), 31-49.

Winsler, A., Hutchison, L. A., De Feyter, J. J., Manfra, L., Bleiker, C., Hartman, S. C., & Levitt, J. (2012). Child, Family, and Childcare Predictors of Delayed School Entry and Kindergarten Retention among Linguistically and Ethnically Diverse Children. Developmental Psychology, 48(5), 1299-1314.

Witmer, S. M., Hoffman, L. M., & Nottis, K. E. (2004). Elementary Teachers' Beliefs and Knowledge about Grade Retention: How Do We Know What They Know?. Education, 125(2), 173.

Wright, W. E., & Arizona State University, E. U. (2005). Evolution of Federal Policy and Implications of No Child Left Behind for Language Minority Students. Policy Brief. Education Policy Research Unit,

Wright, W. E., & Li, X. (2008). High-Stakes Math Tests: How "No Child Left Behind" Leaves Newcomer English Language Learners behind. Language Policy, 7(3), 237-266.

Wu, W., West, S. G., & Hughes, J. N. (2008). Effect of Retention in First Grade on Children's Achievement Trajectories over 4 Years: A Piecewise Growth Analysis Using Propensity Score Matching. Journal Of Educational Psychology, 100(4), 727-740.

Wu, W., West, S. G., & Hughes, J. N. (2008). Short-term effects of grade retention on the growth rate of Woodcock–Johnson III broad math and reading scores. Journal of School Psychology, 46(1), 85-105.

Wu, W., West, S. G., & Hughes, J. N. (2010). Effect of grade retention in first grade on psychosocial outcomes. Journal of Educational Psychology, 102(1), 135.

Xia, C., Glennie, E., & Duke Univ., D. y. (2005). Grade Retention: A Three Part Series. Policy Briefs. Grade Retention: A Flawed Education Strategy [and] Cost-Benefit Analysis of Grade Retention [and] Grade Retention: The Gap between Research and Practice. Center For Child And Family Policy, Duke University.

Xia, N., & Kirby, S. N. (2009). Retaining students in grade: A literature review of the effects of retention on students' academic and nonacademic outcomes. Retrieved from http://www.rand.org/pubs/technical_reports/TR678

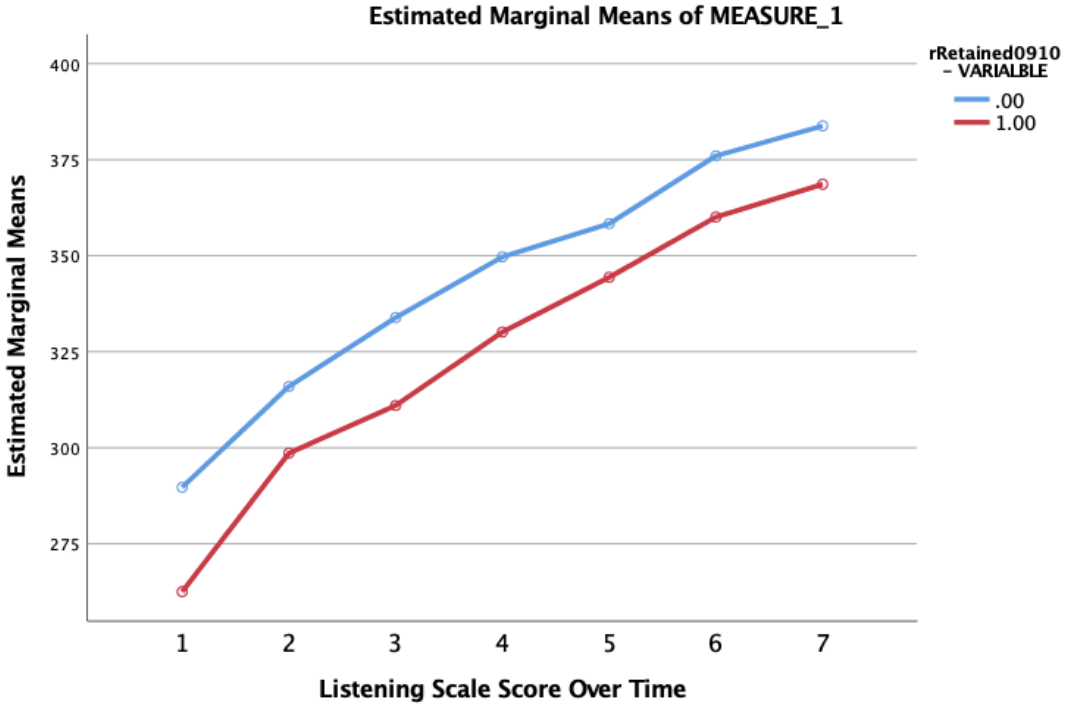
Yanosky, T., Amos, M., Louguit, M., Olson, Cameron, C., Louguit, M., MacGregor, D., Yen, S., and Kenyon, D.M. (2013). Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test, Series 203, 2011-2012 Administration (WIDA Consortium Annual Technical Report No. 8).

Yoshikawa, H., & Kalil, A. (2011). The effects of parental undocumented status on the developmental contexts of young children in immigrant families. Child Development Perspectives, 5(4), 291-297.

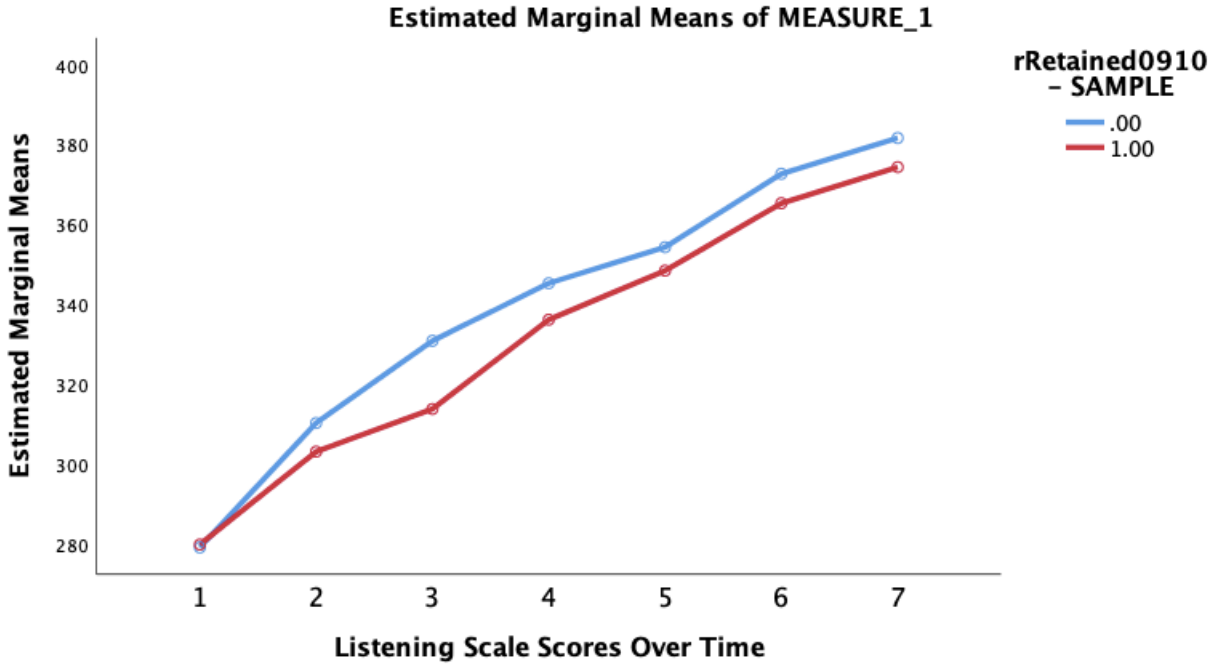
Yoshikawa, H., & Kalil, A. (2011). The effects of parental undocumented status on the developmental contexts of young children in immigrant families. Child Development Perspectives, 5(4), 291-297.

Appendix A. Repeated Measures ANOVA

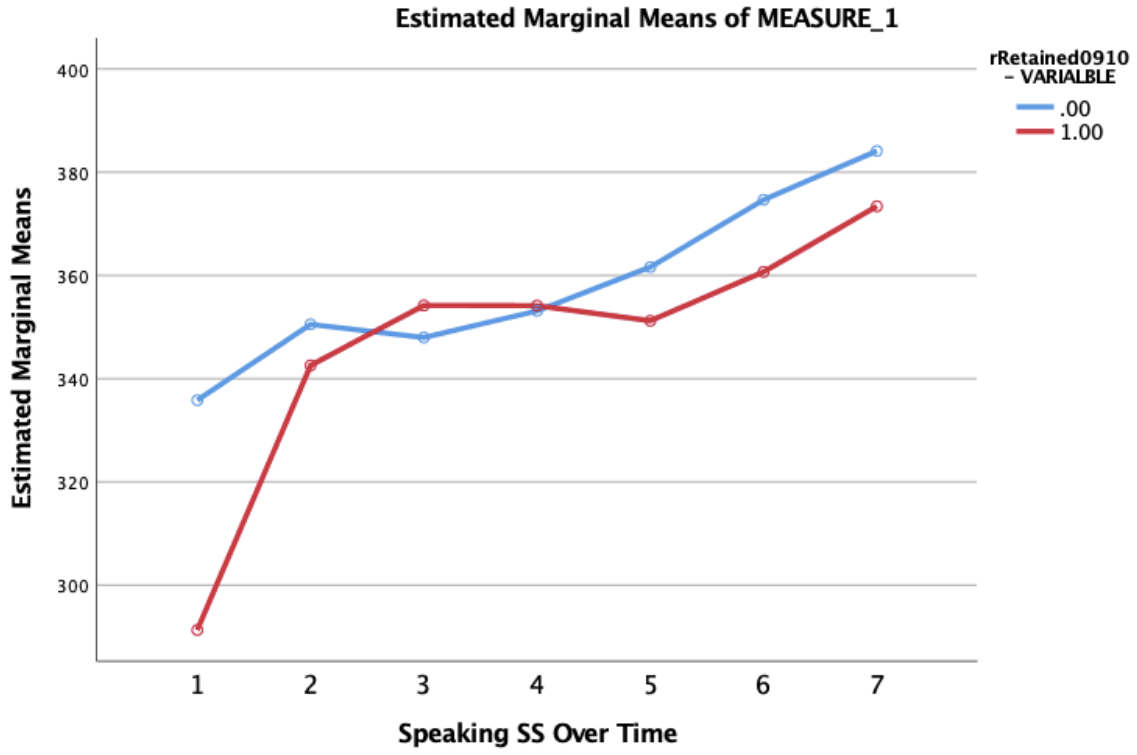
Listening – Unweighted



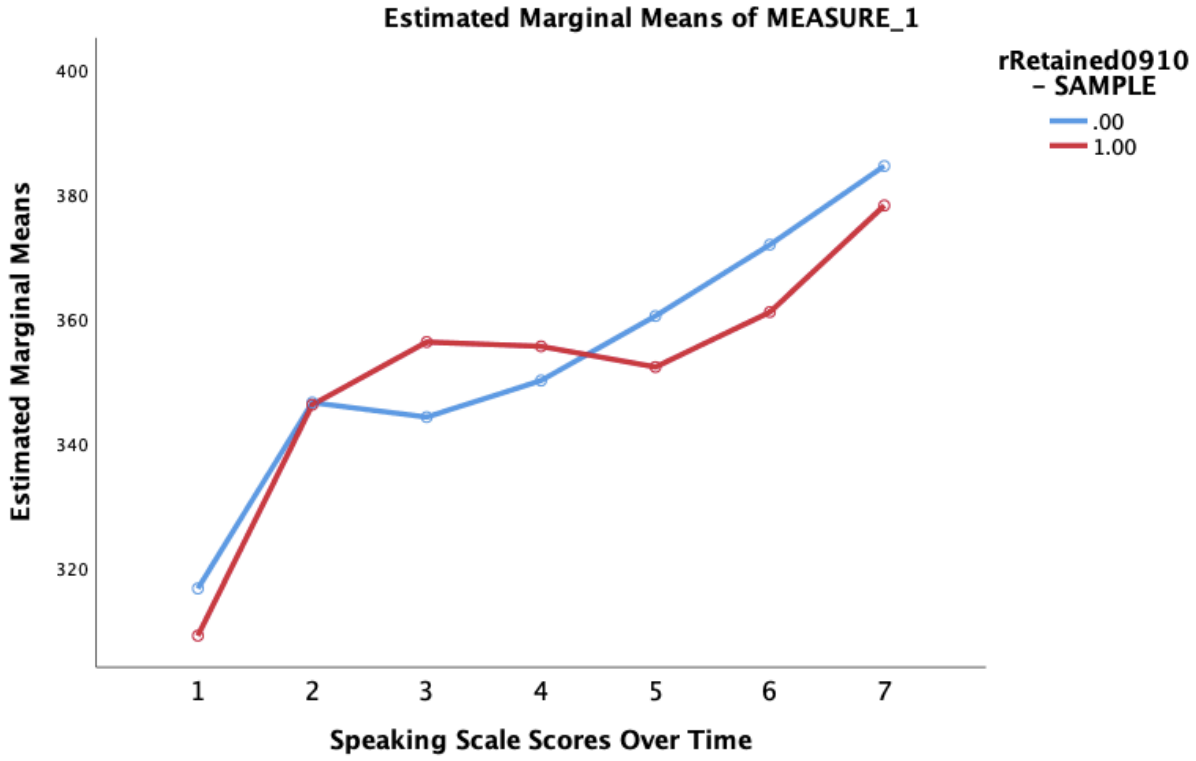
Listening – Weighted



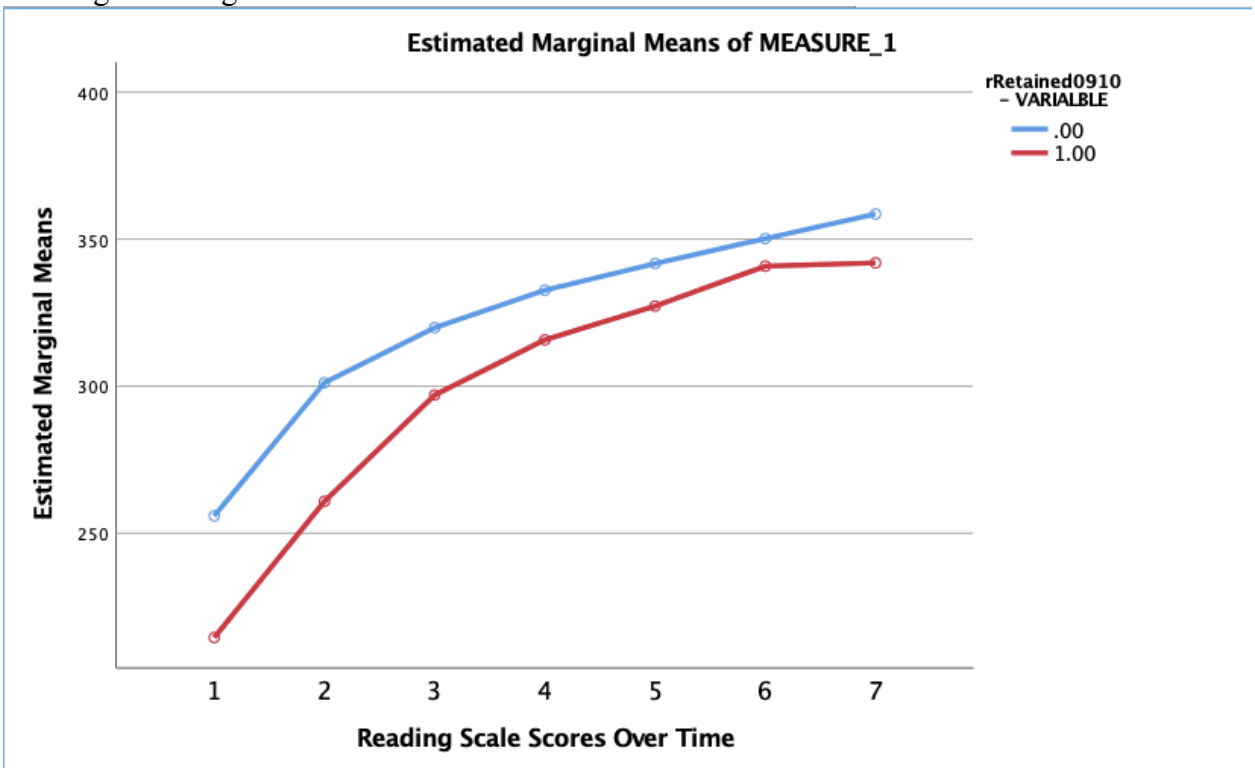
Speaking - Unweighted



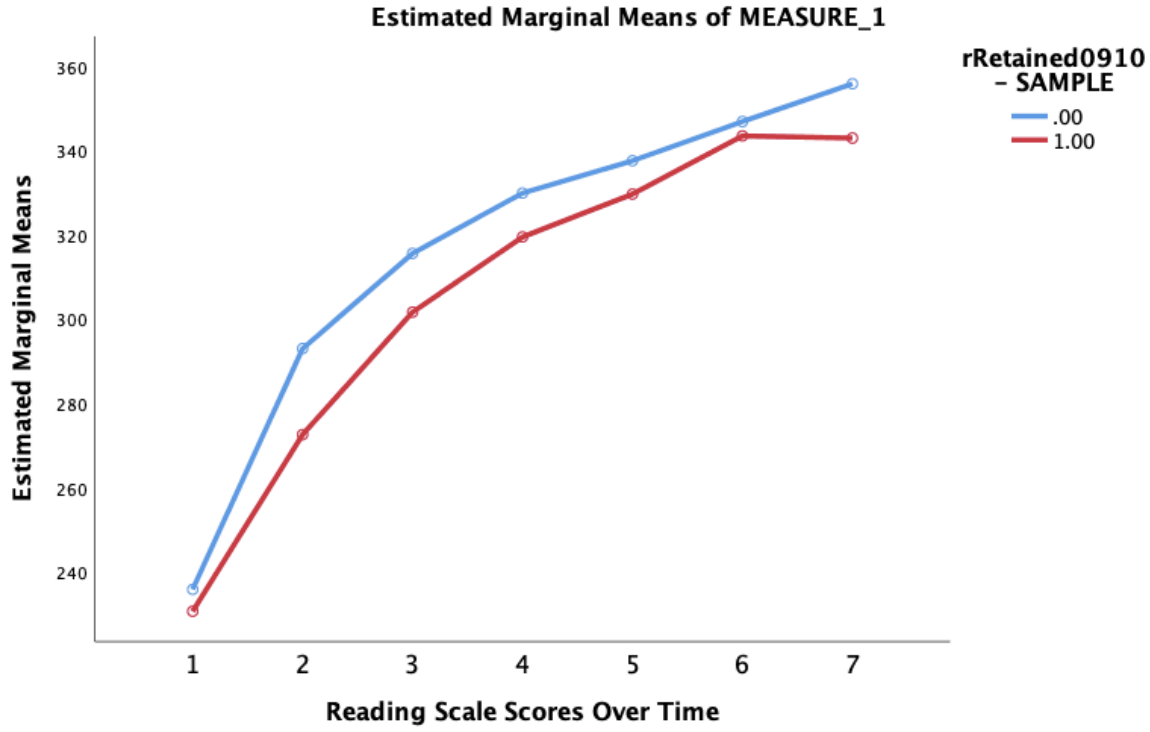
Speaking – Weighted



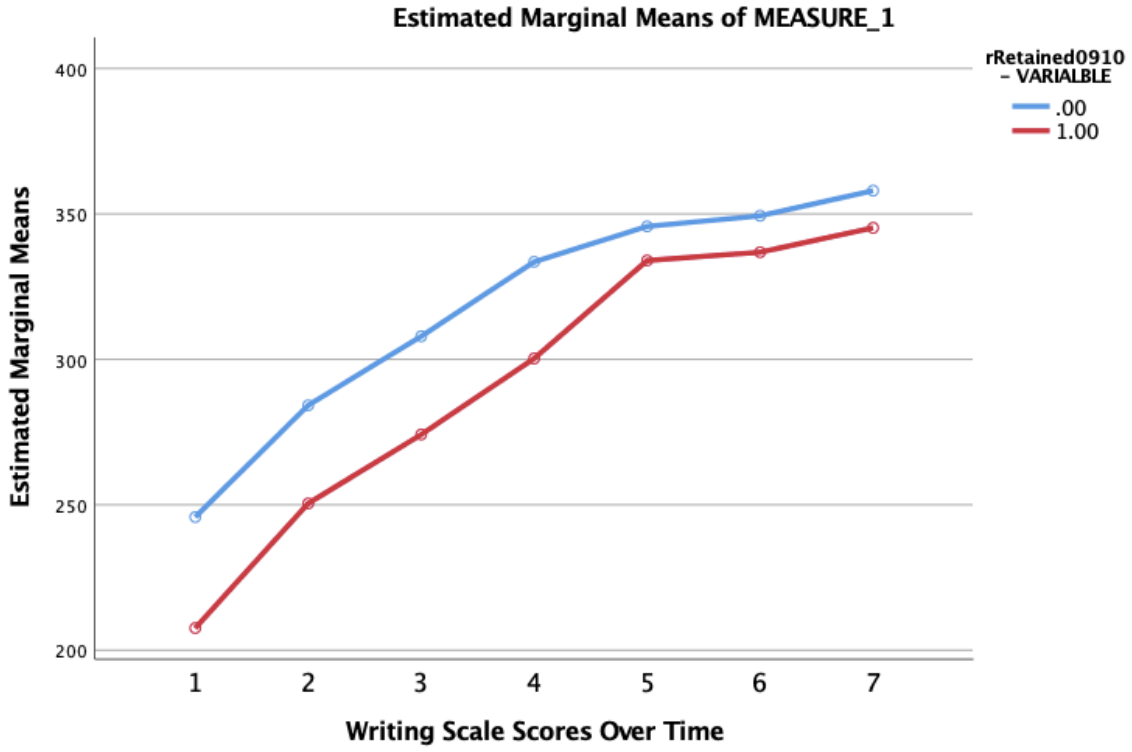
Reading - Unweighted



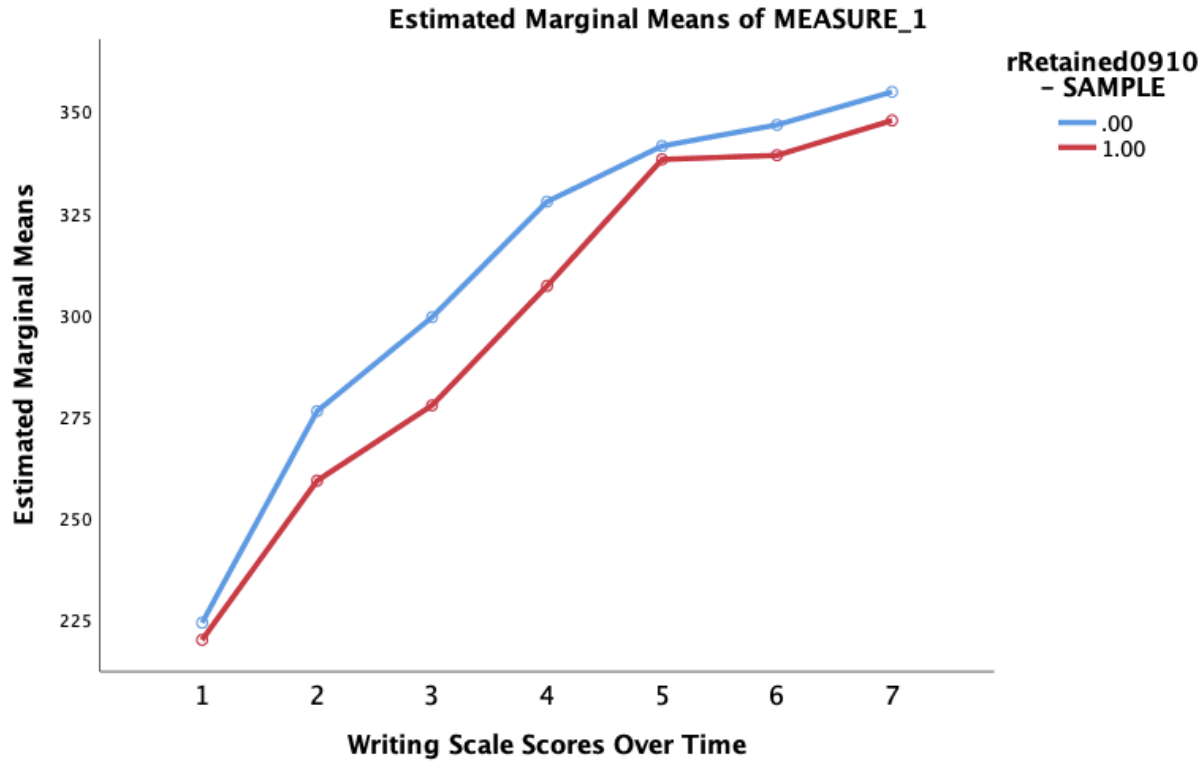
Reading – Weighted



Writing - Unweighted



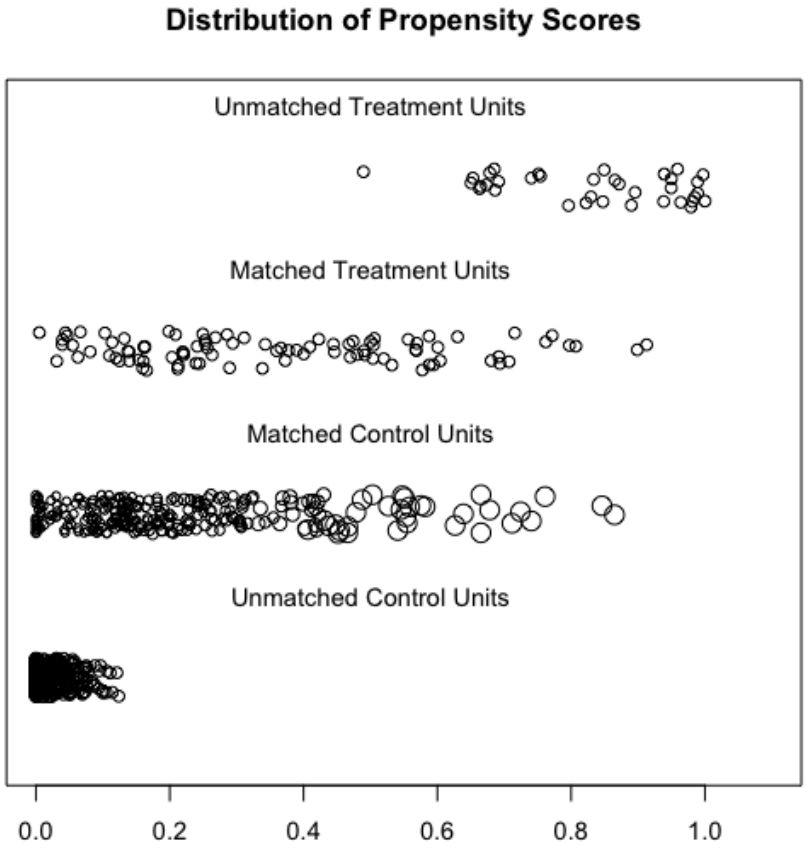
Writing - Weighted

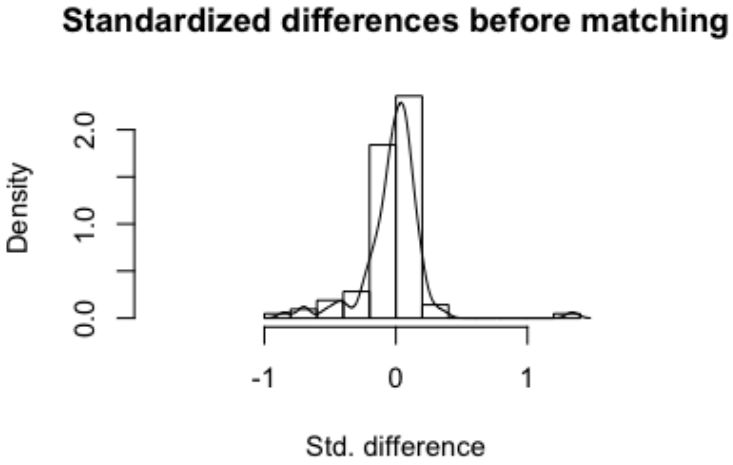


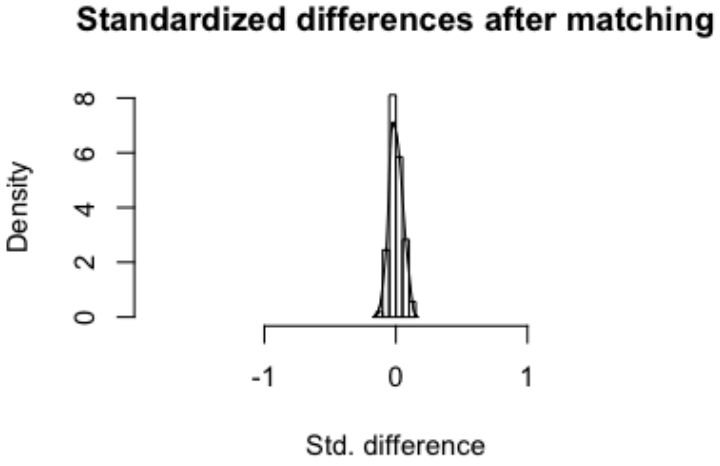
Appendix B. Propensity Score Matching

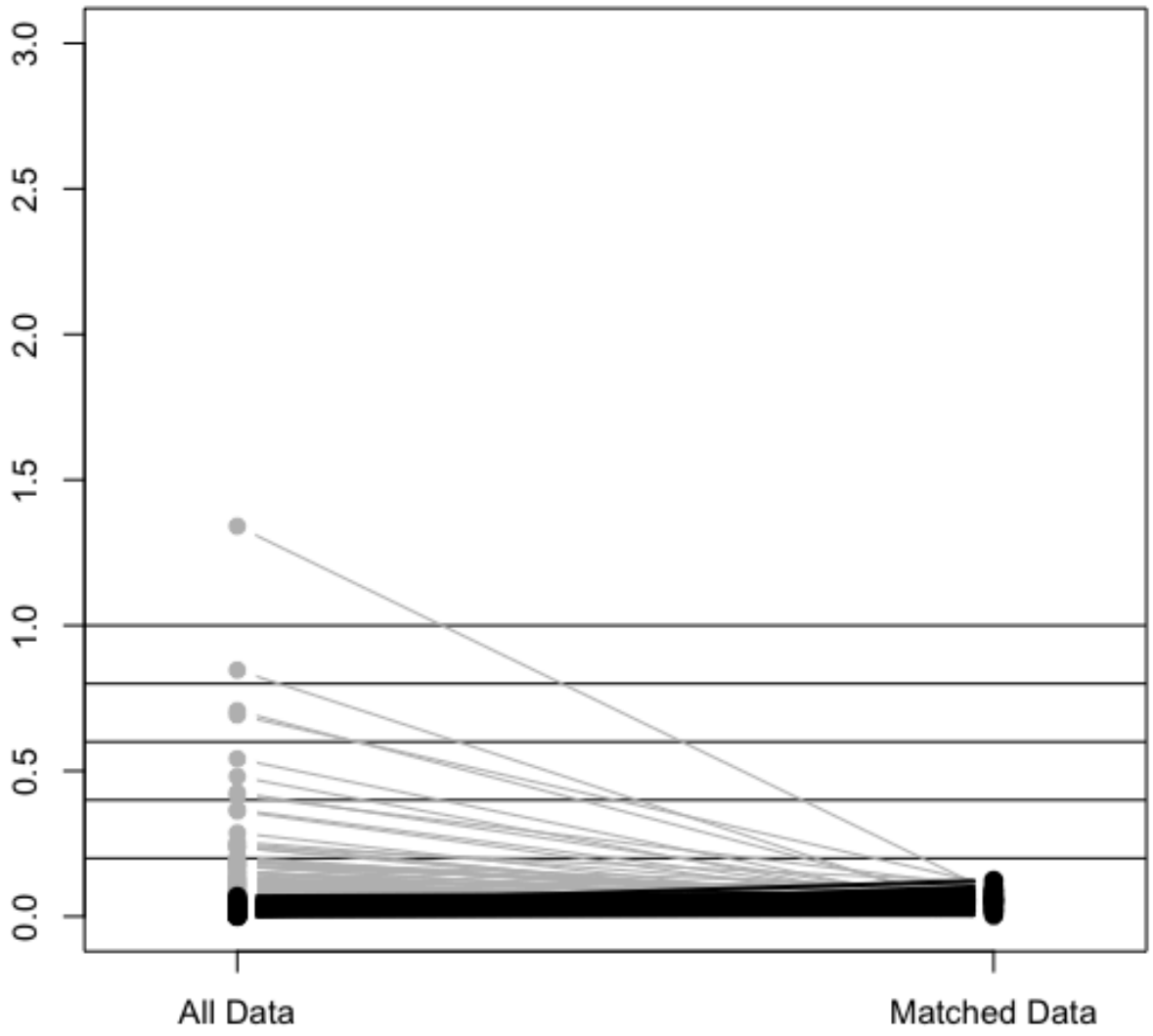
Table B.1
Covariate Descriptive Statistics Before Matching

	Valid	**Mean**	**Min**	**Max**
Initial Grade	828	.923	0	2
Special Education Status	828	.07	0	1
Age is Typical for Grade	828	.731	0	1
Absenteeism (≥ 18)	828	.057	0	1
Language (Spanish)	828	.987	0	1
Ethnicity				
Asian	828	.011	0	1
African American	828	.004	0	1
Native American	828	.005	0	1
White	828	.006	0	1
Gender (Male)	828	.55	0	1
Free/Reduced Lunch	828	.675	0	1
Born Outside of US	828	.108	0	1
Listening PL	828	4.010	1.0	6.0
Speaking PL	828	3.842	1.0	6.0
Reading PL	828	3.153	1.0	6.0
Writing PL	828	2,411	1.0	4.4


Distribution of Propensity Scores
Unmatched Treatment Units
Matched Treatment Units
Matched Control Units
Unmatched Control Units
0.0
0.2
0.4
0.6
0.8
1.0

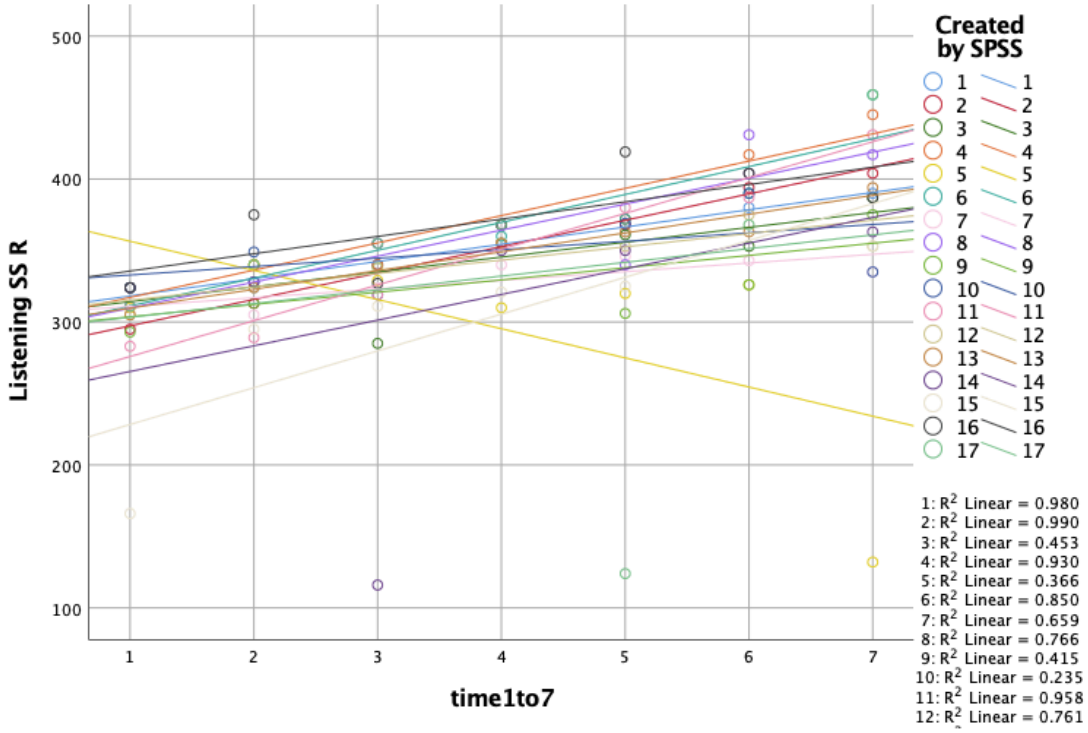

Standardized differences before matching
Density
0.0
1.0
2.0
-1
0
1
Std. difference

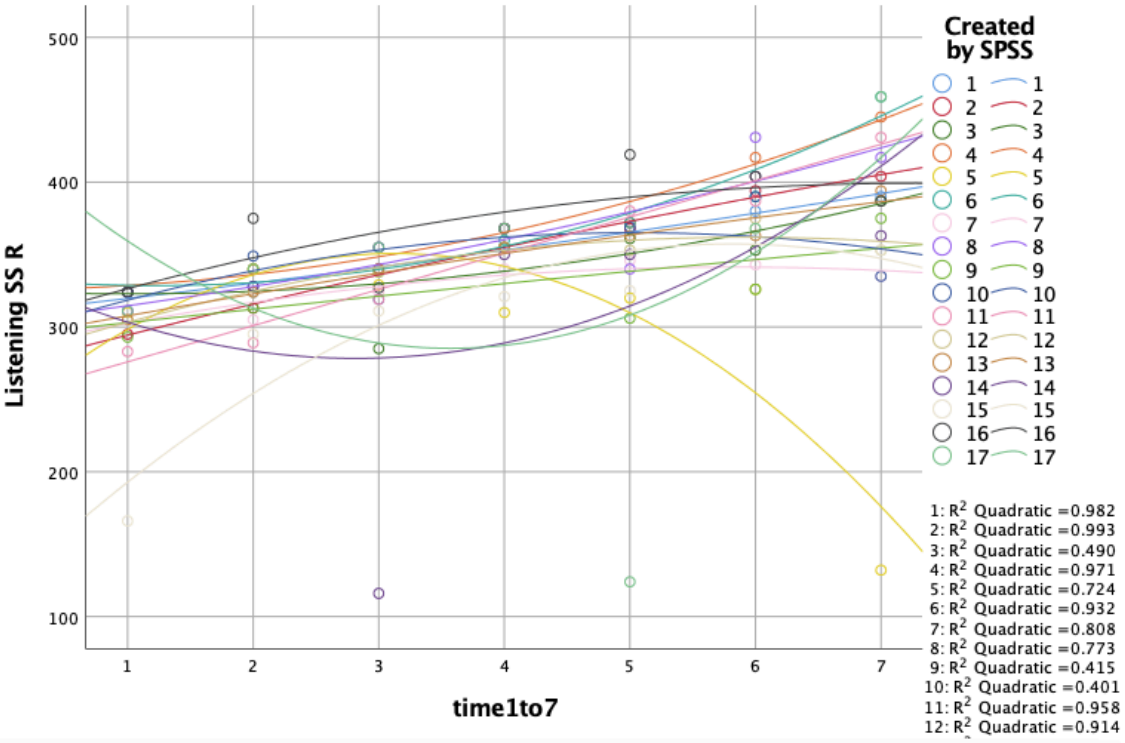

Standardized differences after matching
Density
0
2
4
6
8
-1
0
1
Std. difference


3.0
2.5
2.0
1.5
1.0
0.5
0.0
All Data
Matched Data

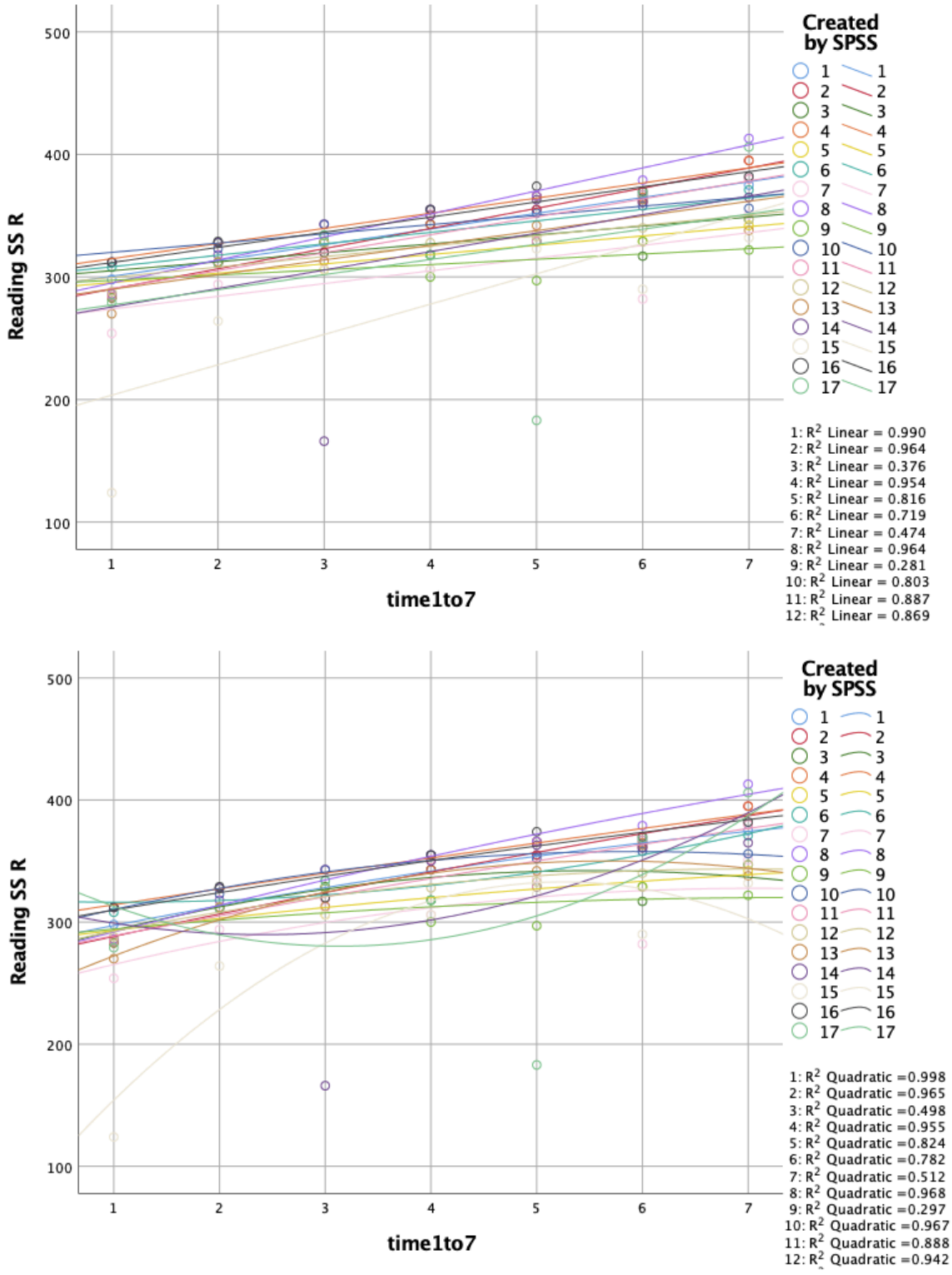
Appendix C. Linear Trends

Listening Scale Score Linear Trajectories – Linear and Quadratic

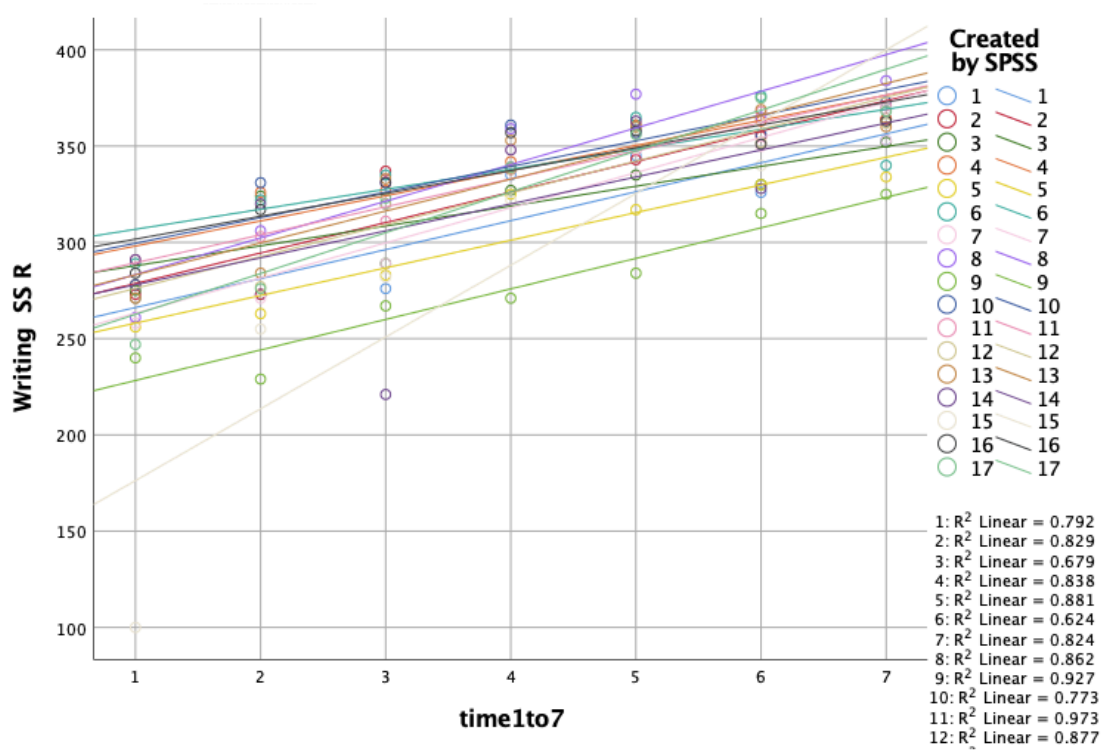


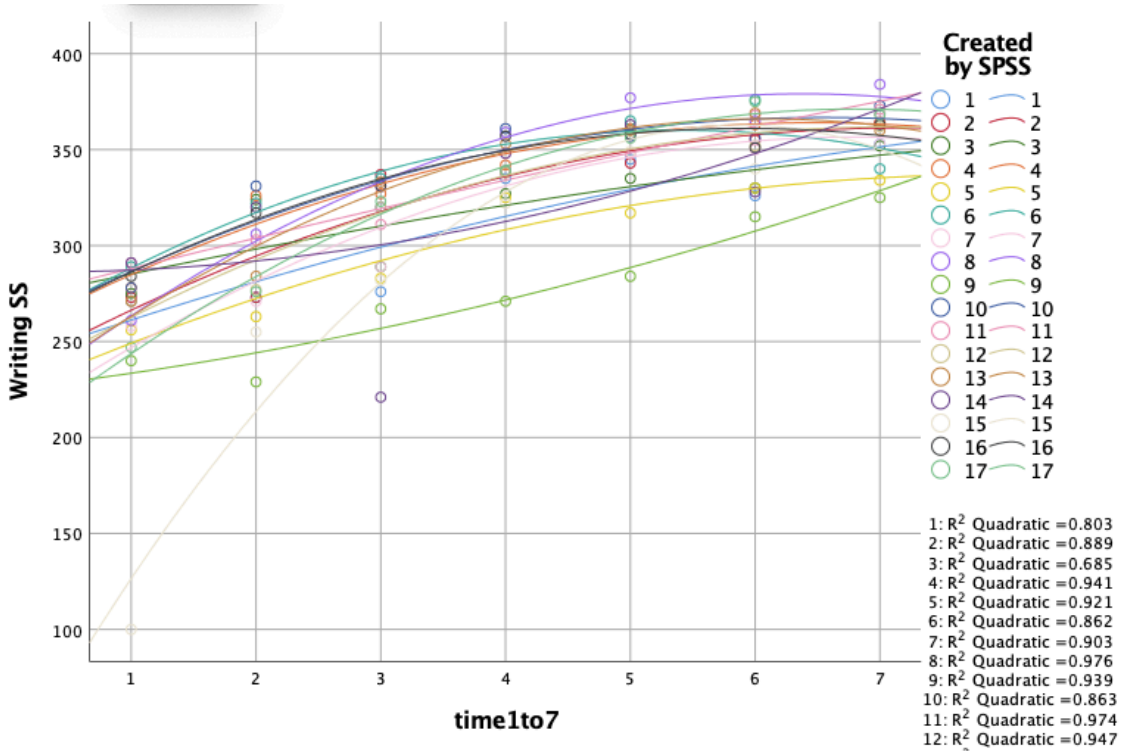


Reading Scale Score Linear Trajectories – Linear and Quadratic

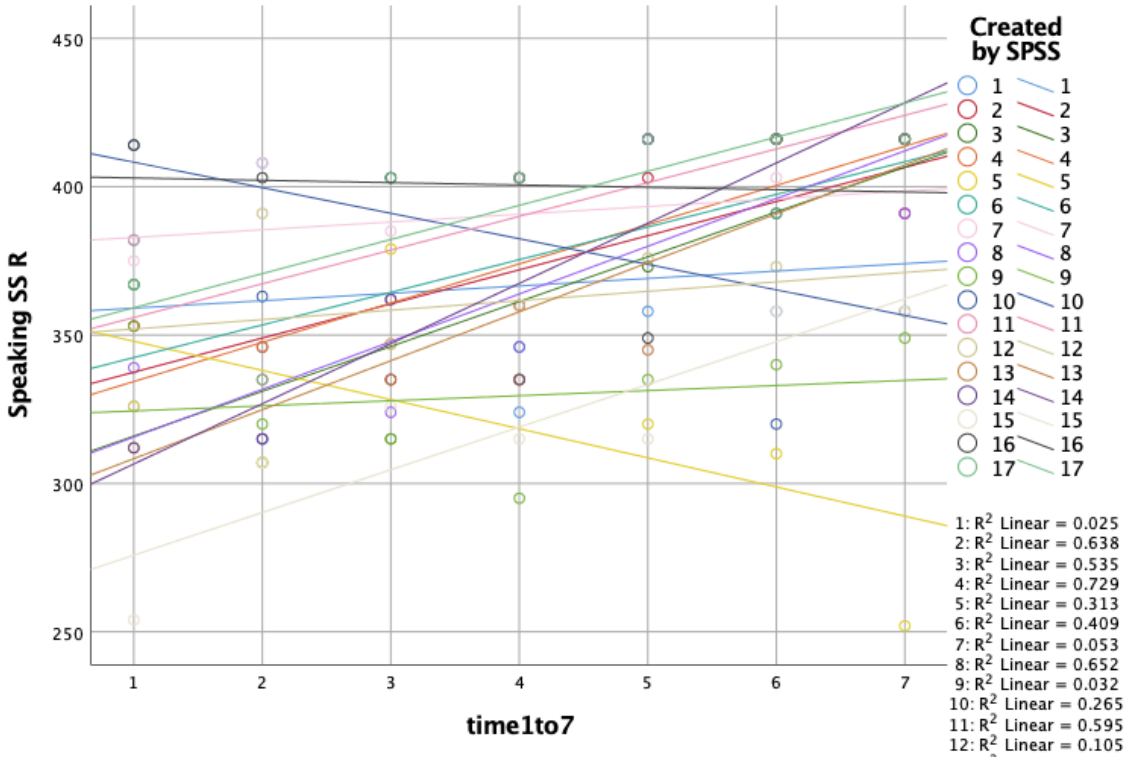


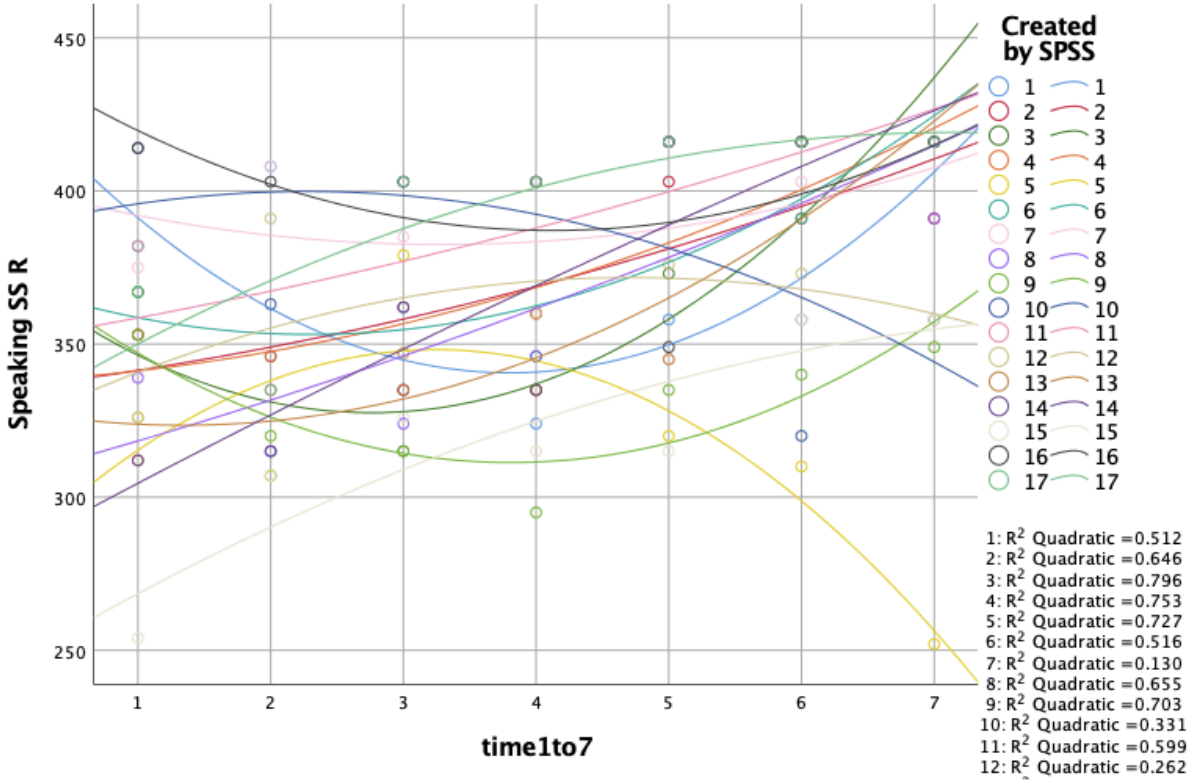
Writing Scale Score Linear Trajectories – Linear and Quadratic


Writing SS R
time1to7
Created by SPSS
1: R² Linear = 0.792
2: R² Linear = 0.829
3: R² Linear = 0.679
4: R² Linear = 0.838
5: R² Linear = 0.881
6: R² Linear = 0.624
7: R² Linear = 0.824
8: R² Linear = 0.862
9: R² Linear = 0.927
10: R² Linear = 0.773
11: R² Linear = 0.973
12: R² Linear = 0.877


Writing SS
time1to7
Created by SPSS
1: R² Quadratic =0.803
2: R² Quadratic =0.889
3: R² Quadratic =0.685
4: R² Quadratic =0.941
5: R² Quadratic =0.921
6: R² Quadratic =0.862
7: R² Quadratic =0.903
8: R² Quadratic =0.976
9: R² Quadratic =0.939
10: R² Quadratic =0.863
11: R² Quadratic =0.974
12: R² Quadratic =0.947

Speaking Scale Score Linear Trajectories – Linear and Quadratic (variable eliminated)





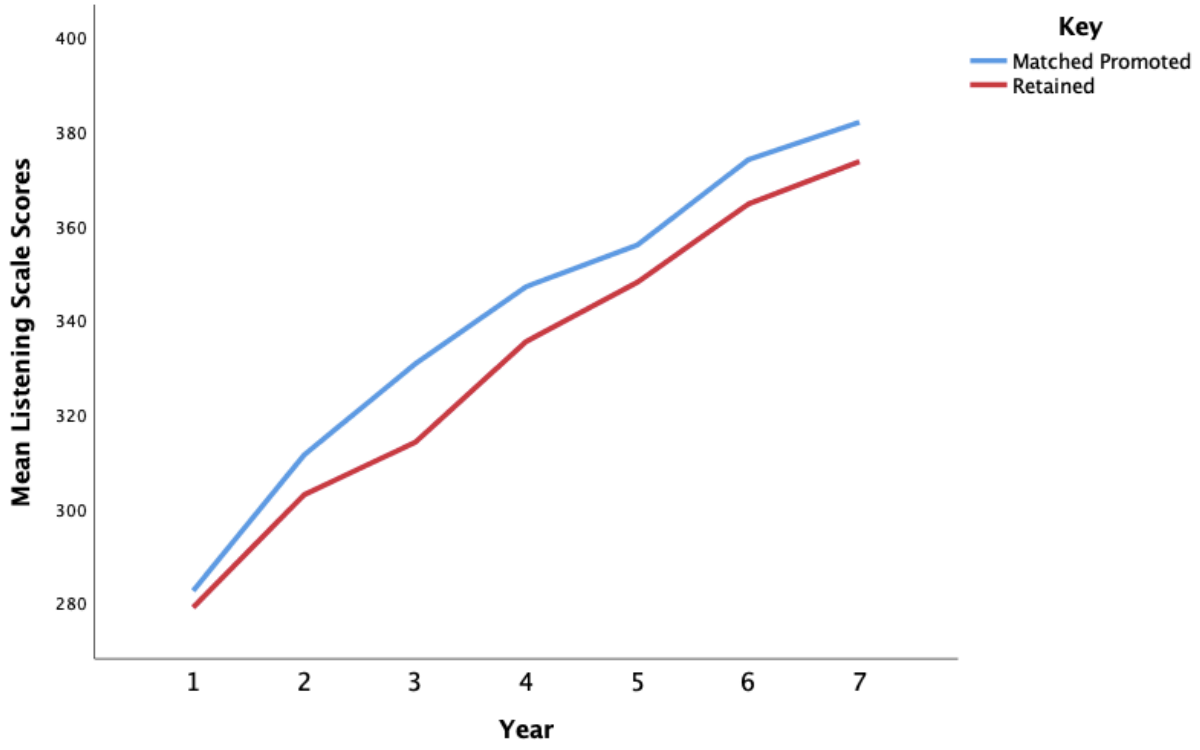
Appendix D: Input Instructions for MPlus

```
Title: Model 2 Listening;

Data: File is C:\Users\Angela Urick\Desktop\ELReten.dat;
!Format is;

Variable: Names are ID SchID Rank ListenSS SpeakSS ReadSS WriteSS
T06 quadT orthquad orthoT Retai910 psweight
IniGradK Sped09 AgeAvgGd Absent Black EtNA White
GenderM LunchFR ForeignB ListenPL SpeakPL ReadPL WritePL
MetBench Hisp50 Mobility TExp;

Usevariables are ID ListenSS orthoT psweight Retai910 IniGradK Sped09 AgeAvgGd
Absent GenderM LunchFR ForeignB ListenPL WritePL
MetBench Hisp50 Mobility TExp;

Cluster = ID;

Within = orthoT;

between = Retai910 IniGradK Sped09 AgeAvgGd Absent
GenderM LunchFR ForeignB ListenPL WritePL
MetBench Hisp50 Mobility TExp;

weight = psweight;

Analysis: Type = twolevel random;

Estimator is MLR;

Model: %between%
ListenSS on Retai910 IniGradK Sped09 AgeAvgGd Absent
GenderM LunchFR ForeignB ListenPL WritePL
MetBench Hisp50 Mobility TExp;

S on Retai910 IniGradK Sped09 AgeAvgGd Absent
GenderM LunchFR ForeignB ListenPL WritePL
MetBench Hisp50 Mobility TExp;

ListenSS with S;

%within%
S | ListenSS on OrthoT;
Output: sampstat TECH1;
```

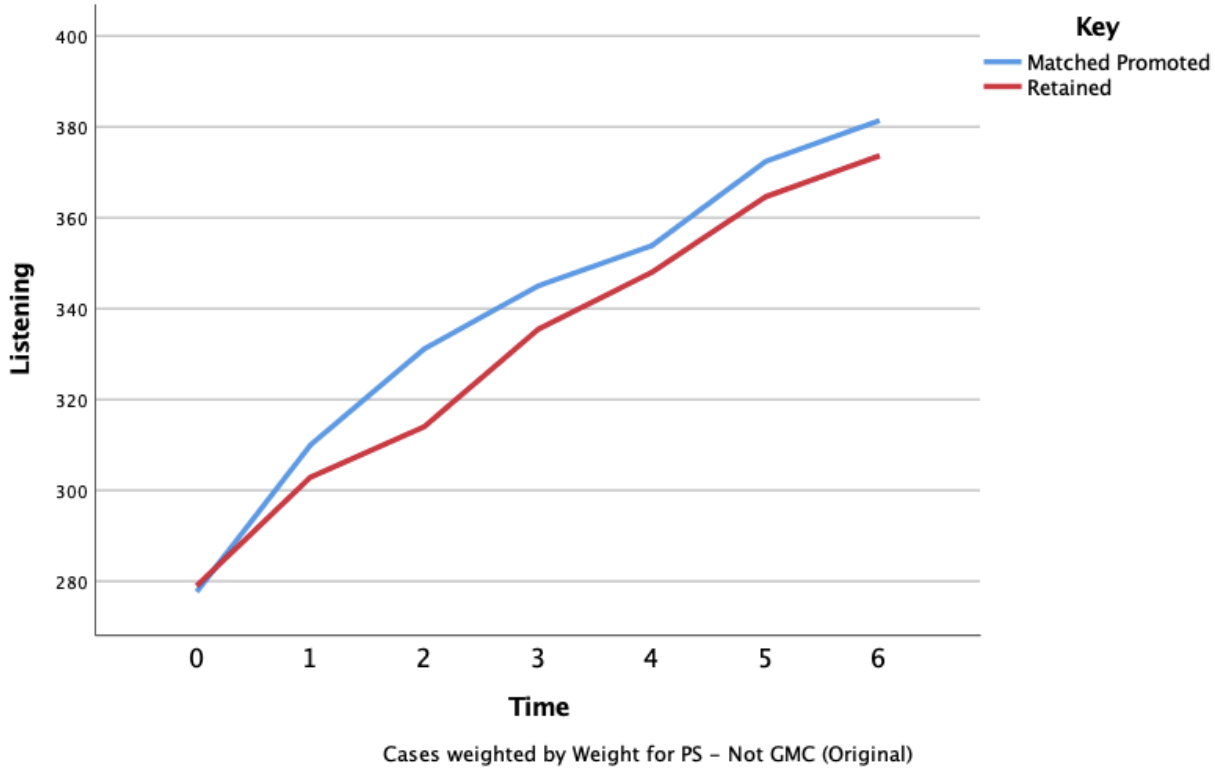
```
Title: Model 2 Reading;

Data: File is C:\Users\Angela Urick\Desktop\ELReten.dat;
!Format is;

Variable: Names are ID SchID Rank ListenSS SpeakSS ReadSS WriteSS
T06 quadT orthquad orthoT Retai910 psweight
IniGradK Sped09 AgeAvgGd Absent Black EtNA White
GenderM LunchFR ForeignB ListenPL SpeakPL ReadPL WritePL
MetBench Hisp50 Mobility TExp;

Usevariables are ID ReadSS orthoT psweight Retai910 IniGradK Sped09 AgeAvgGd
Absent GenderM LunchFR ForeignB ListenPL WritePL
MetBench Hisp50 Mobility TExp;

Cluster = ID;

Within = orthoT;
between = Retai910 IniGradK Sped09 AgeAvgGd Absent
GenderM LunchFR ForeignB ListenPL WritePL
MetBench Hisp50 Mobility TExp;

weight = psweight;

Analysis: Type = twolevel random;

Estimator is MLR;

Model: %between%
ReadSS on Retai910 IniGradK Sped09 AgeAvgGd Absent
GenderM LunchFR ForeignB ListenPL WritePL
MetBench Hisp50 Mobility TExp;

S on Retai910 IniGradK Sped09 AgeAvgGd Absent
GenderM LunchFR ForeignB ListenPL WritePL
MetBench Hisp50 Mobility TExp;

ReadSS with S;

%within%
S | ReadSS on OrthoT;

Output: sampstat TECH1;
```

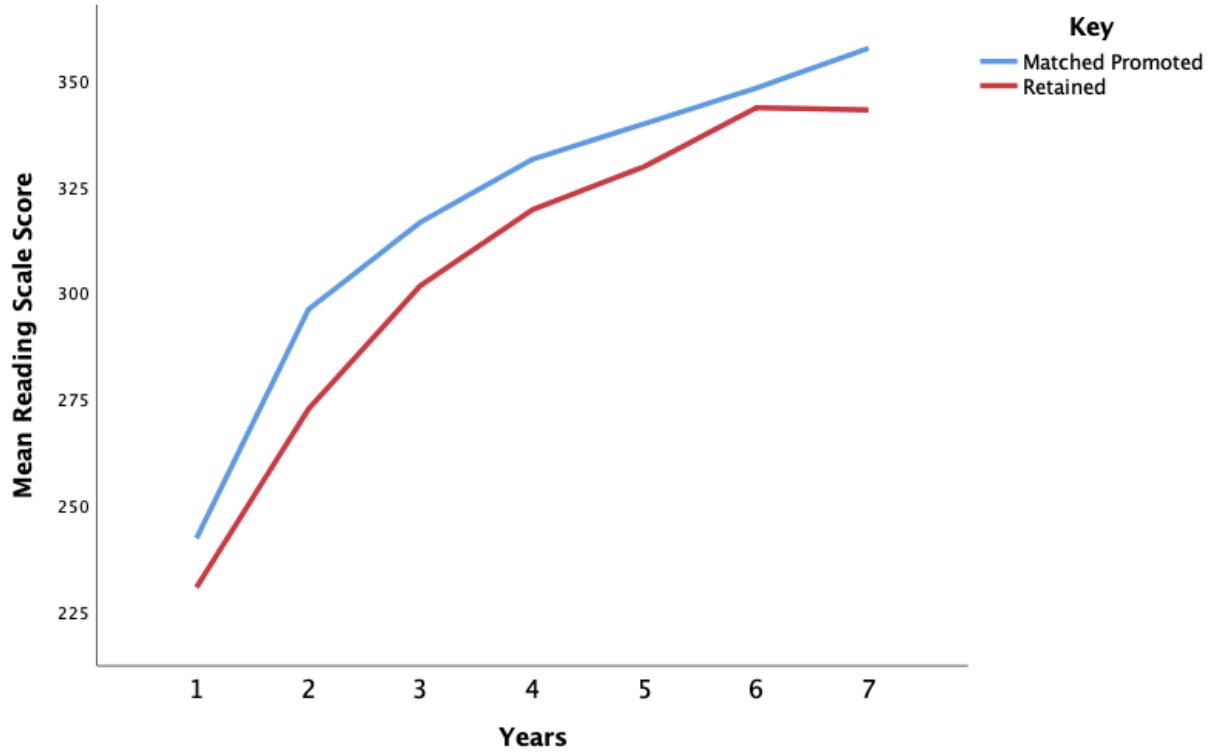
```
Title: Model 2 writing;

Data: File is C:\Users\Angela Urick\Desktop\ELReten.dat;
!Format is;

Variable: Names are ID SchID Rank ListenSS SpeakSS ReadSS WriteSS
T06 quadT orthquad orthoT Retai910 psweight
IniGradK Sped09 AgeAvgGd Absent Black EtNA White
GenderM LunchFR ForeignB ListenPL SpeakPL ReadPL WritePL
MetBench Hisp50 Mobility TExp;

Usevariables are ID WriteSS orthoT psweight Retai910 IniGradK Sped09 AgeAvgGd
Absent GenderM LunchFR ForeignB ListenPL WritePL
MetBench Hisp50 Mobility TExp;

Cluster = ID;

Within = orthoT;

between = Retai910 IniGradK Sped09 AgeAvgGd Absent
GenderM LunchFR ForeignB ListenPL WritePL
MetBench Hisp50 Mobility TExp;

weight = psweight;

Analysis: Type = twolevel random;

Estimator is MLR;

Model: %between%
WriteSS on Retai910 IniGradK Sped09 AgeAvgGd Absent
GenderM LunchFR ForeignB ListenPL WritePL
MetBench Hisp50 Mobility TExp;

S on Retai910 IniGradK Sped09 AgeAvgGd Absent
GenderM LunchFR ForeignB ListenPL WritePL
MetBench Hisp50 Mobility TExp;

WriteSS with S;
%within%

S | WriteSS on OrthoT;

Output: sampstat TECH1;
```

Appendix E. Results – Unweighted vs Weighted

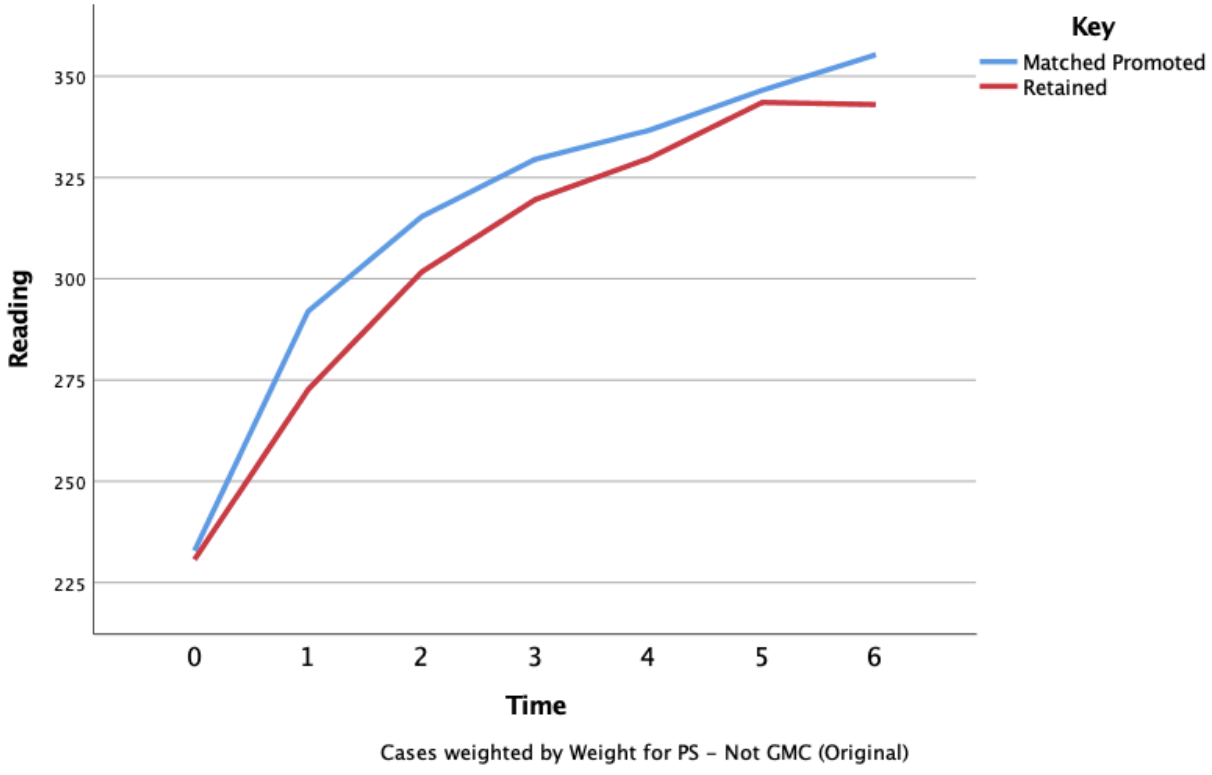


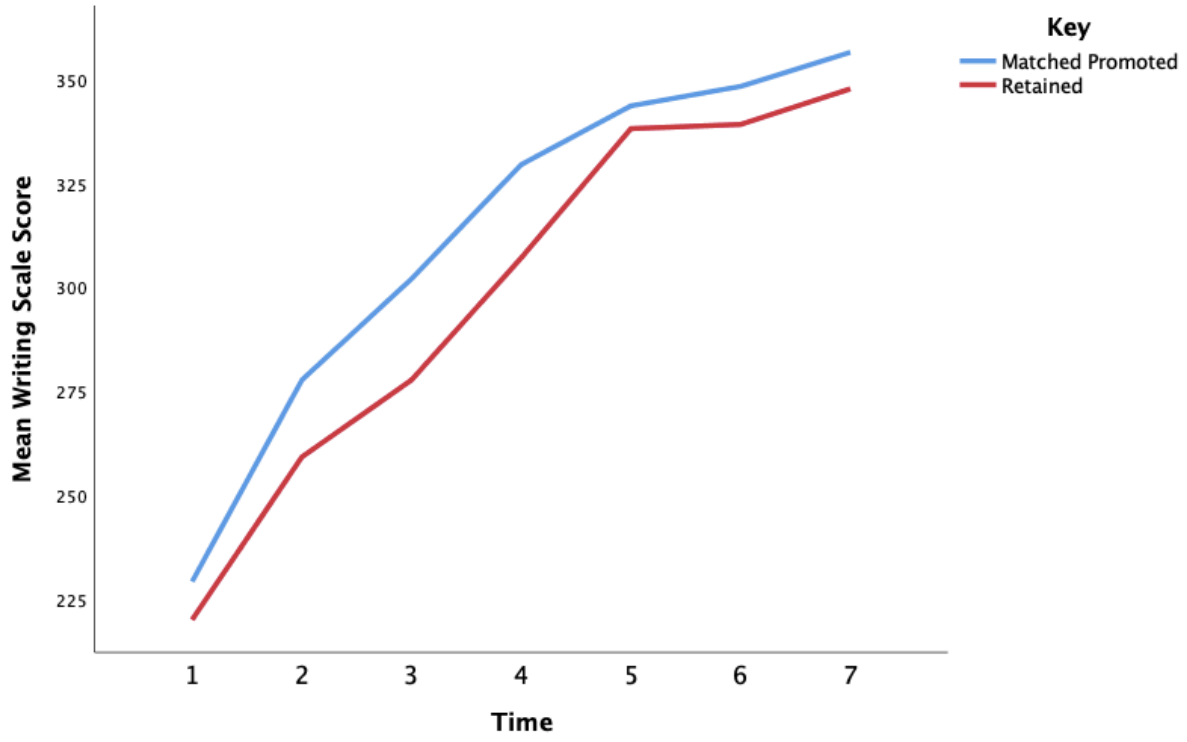
Weighted



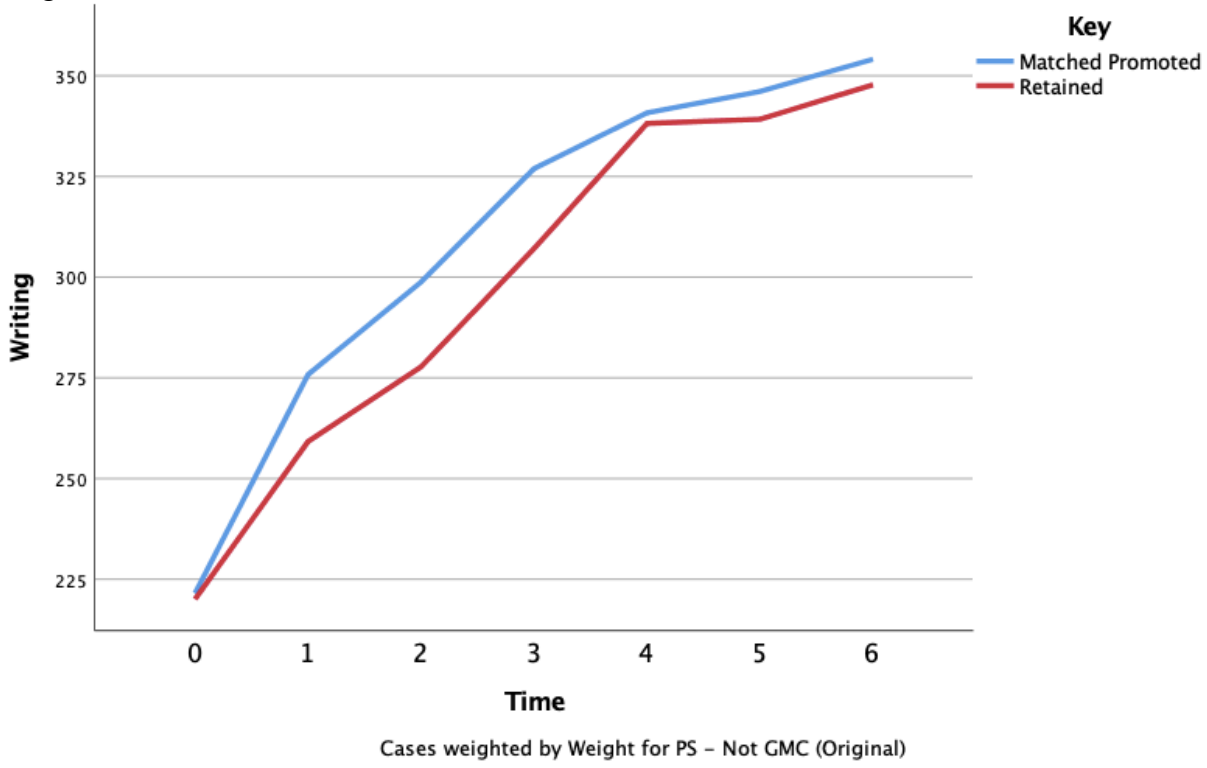

Key
Matched Promoted
Retained
Mean Reading Scale Score
350
325
300
275
250
225
1
2
3
4
5
6
7
Years

Weighted


Key
Matched Promoted
Retained
Reading
350
325
300
275
250
225
0
1
2
3
4
5
6
Time
Cases weighted by Weight for PS – Not GMC (Original)


Key
Matched Promoted
Retained
Mean Writing Scale Score
350
325
300
275
250
225
1
2
3
4
5
6
7
Time

Weighted


Key
Matched Promoted
Retained
Writing
350
325
300
275
250
225
0
1
2
3
4
5
6
Time
Cases weighted by Weight for PS – Not GMC (Original)