

The Treatment Effect of the City Connects Intervention on Exiting Limited English Proficiency Status:

Author: Bercem Akbayin

Persistent link: <http://hdl.handle.net/2345/bc-ir:107654>

This work is posted on [eScholarship@BC](#),
Boston College University Libraries.

Boston College Electronic Thesis or Dissertation, 2017

Copyright is held by the author, with all rights reserved, unless otherwise noted.

Boston College
Lynch School of Education

Department of
Measurement, Evaluation, Statistics, and Assessment

THE TREATMENT EFFECT OF THE CITY CONNECTS INTERVENTION ON EXITING
LIMITED ENGLISH PROFICIENCY STATUS

Dissertation
by

BERCEM AKBAYIN

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

December 2017

THE TREATMENT EFFECT OF THE CITY CONNECTS INTERVENTION ON EXITING
LIMITED ENGLISH PROFICIENCY STATUS

by

Bercem Akbayin, Author

Dr. Henry I. Braun, Dissertation Chair

ABSTRACT

The *City Connects* intervention is motivated by the belief that out-of-school factors act as barriers to student thriving in cognitive and non-cognitive domains. It seeks to address these barriers first by identifying each student's strengths and needs and then by providing a tailored set of prevention, intervention, and enrichment programs. Underlying the program is the assumption that provision of high-quality resources and individualized services will enable children to be cognitively, socio-emotionally, and physically prepared to thrive in school.

This study's purpose was to estimate the effects of the *City Connects* intervention on English learners' (EL) likelihood of exiting Limited English Proficiency (LEP) status. ELs comprise a student subpopulation most at-risk to fail academically, and exposure to the program was hypothesized to improve their likelihood of exiting LEP status earlier than otherwise. A series of one- and two-level discrete-time event history analyses were conducted on the main analytic sample as well as two sub-samples. As participation in *City Connects* is at the school-level, school-level matching was used for sub-samples 1 and 2, and propensity score weights were applied at the student-level for all three samples. Additionally, hazard probabilities, survival probabilities, cumulative hazard rates, and median lifetimes were estimated. Lastly, a

sensitivity analysis was conducted to examine whether effects were robust to unobserved selection bias.

The results indicated that ELs participating in the *City Connects* intervention were significantly more likely to exit LEP status earlier than their peers in comparison schools. The median time in LEP status in *City Connects* schools was shorter and translated into a gain of at least one half of a year in grade in mainstream classes. Also, all the fitted models indicated that approximately 10 percent more *City Connects* students exited LEP status by the end of fifth grade than comparison students. Findings highlight the impact of the *City Connects* intervention, as ELs entering mainstream classes earlier could translate into important academic and non-academic gains, such as improved academic achievement and increased self-confidence.

ACKNOWLEDGEMENTS

Many people have helped to shape and define the person I am today. Among those who made it possible for me to complete this dissertation, I'd like to thank, first and foremost, my dissertation advisor, Dr. Henry Braun. His guidance on the technical aspects of this work and his careful reading of my drafts were invaluable. He always encouraged me to look at the big picture and make meaningful connections between the methods and the real-life implications of this research.

I'd also like to extend special thanks to my readers, Dr. Laura O'Dwyer and Dr. Eric Dearing. Thank you, Dr. O'Dwyer, for giving me technical advice, for teaching me advanced methodology along my journey in the MESA department, and for providing detailed feedback on the dissertation drafts. I'm also grateful to you for acting as a female role model in a field where they are rare. Dr. Dearing, thank you for your insights from the child development and poverty fields. Your thoughtful questions and constructive feedback were essential for the successful completion of this dissertation.

I'd also like to extend my thanks to those associated with the *City Connects* for their help and support, without which this research would not have been possible. Dr. Mary Walsh, the principal investigator of the *City Connects* intervention, I am grateful for being able to use the program's data in my dissertation, and Anastasia Raczek, my advisor and mentor at *City Connects*, introduced me to many exciting projects there. My time with *City Connects* gave me the opportunity to put everything I had learned in my graduate study program into practice as well as challenging myself with new techniques. I'd also like to thank Dr. Terrence J. Lee-St. John for his support both as a friend and colleague during my time at MESA and at *City Connects*. Thanks for being the "go-to person" whenever I needed help and for always

welcoming my questions and making time to think through the answers with me. In short, my time with *City Connects* was invaluable and provided the greatest learning environment a graduate student could ever hope for.

My parents, Kadriye and Hasan Akbayin, and sister Zelal, this dissertation is dedicated to you. My father taught me the foundations of mathematics, philosophy, and humor. Because of him, I was able to look at the world through the lens of science and find joy in life whatever the circumstances. My mother instilled in me a passion for life and taught me that love and compassion only grow by sharing. Because of her endless love and care, I was always surrounded by a community that loved and supported me through the challenges of life. Thanks to my parents, I grew up knowing that if I fell down, there would always be someone to catch me. And, my sister Zelal, you remind me to meet every day with laughter. Thank you for always lifting me up and reminding us all that everything is and will be ok.

And, lastly, Ozgur Sahin, my dear husband, I will forever ask myself how I got so lucky as to have a partner like you. Your continued enthusiasm, interest, and encouragement have kept me going throughout the long, hard process of completing this research. Thank you for being an inspiration every day of my life.

Table of Contents

Chapter 1. Introduction	1
1.1 <i>City Connects</i> Intervention	6
1.2 Purpose of Study and Research Questions	8
1.3 Significance of the Study	11
1.4 Outline of the Dissertation	12
Chapter 2. Literature Review	14
2.1 Typical Processes for Identification, Placement, and Reclassification of EL Students	14
2.2 Challenges to Reclassifying EL Students into Mainstream Classrooms	19
2.2.1 Quality of ELP assessments has implications for reclassification.	19
2.2.2 Standardized achievement tests not designed for reclassification purposes.	21
2.2.3 Lack of a widely accepted second language acquisition theory.	23
2.2.4 Contradictory NCLB incentives for reclassification of EL students.	25
2.3 Research on Reclassification of EL Students as English Proficient	26
2.4 Summary	42
Chapter 3. Methodology	44
3.1 Research Design	44
3.1.1 Study design	44
3.1.2 Threats to internal validity	45
3.1.2.1 Ambiguous temporal precedence	45
3.1.2.2 Selection	46
3.1.2.3 Maturation	46
3.1.2.4 Regression to the mean	47
3.1.2.5 Attrition	47
3.1.2.6 Testing	48
3.1.2.7 Instrumentation	48
3.1.2.8 History	48
3.1.2.9 Assumptions for causality	49
3.1.2.9.1 SUTVA	50
3.1.2.9.2 Strongly ignorable treatment assignment assumption.	52
3.2 Data Description	52
3.2.1 Data sources	52
3.2.2 Sample	53
3.2.3 Outcome variable	55
3.2.4 Student-Level variables	56

3.2.5 School-Level variables.....	58
3.3 Analytic Strategy	59
3.3.1 Stage one: Preliminary descriptive analyses.....	59
3.3.2 Stage two: Estimation of propensity score weights and school level matching.	60
3.3.3 Stage three: Discrete event history analysis to estimate the <i>City Connects</i> effect.....	65
3.3.3.1 Censoring.....	65
3.3.3.2 Estimating the parameters of the discrete-time hazard model.....	67
3.3.3.3 MLE of hazard model parameters.....	69
3.3.4 Stage four: Modeling strategies for the discrete-time hazard model.	71
3.3.4.1 Research question 1.	71
3.3.4.2 Research question 2.	72
3.3.4.3 Research Question 3.	76
3.3.4.4 Research Question 4.	77
Chapter 4. Results.....	81
4.1 Descriptive Analyses	82
4.2 Research Question One.....	88
4.3 Research Question Two.....	94
4.3.1 Baseline equivalence.....	95
4.3.2 Results from model 1 (one-level discrete-time event history models).	96
4.3.3 Results from model 2 (two-level discrete-time event history models).	114
4.4 Research Question Three.....	126
4.5 Research Question Four.....	129
Chapter 5. Conclusions.....	132
5.1 Summary of Findings.....	132
5.1.1 Research question 1.	132
5.1.2 Research question 2.	133
5.1.3 Research question 3.	137
5.1.4 Research question 4.	139
5.2 Limitations and Future Research	141
5.3 Conclusions and Policy Implications.....	148
References.....	151
Appendix A.....	168
Appendix B.....	170

List of Tables

Table 1-1: <i>Percentage of high school graduates who completed specific STEM courses by year and EL status.</i>	4
Table 3-1: <i>City Connects Longitudinal Data Structure.</i>	54
Table 3-2: <i>Sample Student-Level Data Set.</i>	56
Table 3-3: <i>Description of Student-Level Variables.</i>	57
Table 3-4: <i>Description of School-Level Variables.</i>	59
Table 3-5: <i>The Conditional Probability of Parental Involvement U Given Assignment to a City Connects School.</i>	79
Table 3-6: <i>The 10 Pairs of Conditional Probabilities Used in the Simulation of the Unknown Variable U.</i>	80
Table 4-1: <i>Baseline Student Characteristics by Group Membership for the Big Analytic Sample.</i>	83
Table 4-2: <i>Baseline Student Characteristics by Group Membership for Sub- Sample 1.</i>	85
Table 4-3: <i>Baseline Student Characteristics by Group Membership for Sub- Sample 2.</i>	87
Table 4-4: <i>Life Table for the Big Analytic Sample.</i>	90
Table 4-5: <i>Life Table for Sub-Sample 1.</i>	90
Table 4-6: <i>Life Table for Sub-Sample 2.</i>	91
Table 4-7: <i>Descriptive Summary of Exiting LEP Status and Censored Cases.</i>	92
Table 4-8: <i>Median times for Time to Event and Overall Comparison.</i>	94
Table 4-9: <i>Standardized Bias Statistics Before and After ATT Weighting.</i>	96
Table 4-10: <i>Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Sub-Sample 1, ATT Weighted.</i>	101

Table 4-10 (Continued): <i>Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Sub-Sample 1, ATT Weighted.</i>	102
Table 4-11: <i>Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Sub-Sample 2, ATT Weighted.</i>	106
Table 4-11(Continued): <i>Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Sub-Sample 2, ATT Weighted.</i>	107
Table 4-12: <i>Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Big Analytic Sample, ATT Weighted.</i>	119
Table 4-13: <i>Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Big Analytic Sample, ATT Weighted.</i>	120
Table 4-13 (Continued): <i>Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Big Analytic Sample, ATT Weighted.</i>	121
Table 4-14: <i>Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Big Analytic Sample, ATT Weighted.</i>	122
Table 4-14 (Continued): <i>Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Big Analytic Sample, ATT Weighted.</i>	123
Table 4-15: <i>The Estimated Treatment Effects of the City Connects Intervention after Model 2F Was Adjusted for Simulated U Values and the Fixed Regression Coefficient of 0.8.</i>	131
Table 5-1: <i>Median Lifetimes based on final fitted models for each sample.</i>	138

List of Figures

<i>Figure 1-1</i> . NAEP average scale scores for long-term reading, by EL status.....	2
<i>Figure 1-2</i> . NAEP average scale scores for long-term mathematics, by EL status.....	3
<i>Figure 3-1</i> : Illustration of censoring.....	66
<i>Figure 4-1</i> : Fitted hazard probability for exiting LEP status conditional on student characteristics, sub-sample 1.	109
<i>Figure 4-2</i> : Survival probability conditional on student characteristics, sub-sample 1.	111
<i>Figure 4-3</i> : Fitted hazard probability for exiting LEP status conditional on student characteristics, sub-sample 2.	112
<i>Figure 4-4</i> : Survival probability conditional on student characteristics, sub-sample 2.	113
<i>Figure 4-5</i> : Fitted hazard probability of exiting LEP status conditional on student characteristics, Big Analytic Sample.....	125
<i>Figure 4-6</i> : Survival probability conditional on student characteristics, Big Analytic Sample.	126
<i>Figure 4-7</i> : Fitted trend lines for survival probabilities, Big Analytic Sample.....	129

Chapter 1. Introduction

Despite well-established research attesting to the adverse relationship between poverty and child outcomes in the United States (U.S.) (Brooks-Gunn & Duncan, 1997; Dearing, 2008; Evans, 2004), research on the successful mitigation of the effects of childhood poverty is still ongoing. Unfortunately, more than 16 million American children still live in poverty (U.S. Census Bureau, 2013, 2015), and many are considered at risk for inadequate or unhealthy cognitive, socio-emotional, or physical development due to multiple risk factors including, but not limited to, homelessness, violence, inadequate nutrition, environmental toxins, inequality in and lack of access to institutional resources, such as medical and dental care (Barton & Coley, 2009; Brooks-Gunn & Duncan, 1997; Dearing, 2008; Evans, 2004; Lieten, 2008; Yoshikawa, Aber, & Beardslee, 2012). Among children who live in poverty, English Learners (ELs) constitute one of the fastest growing subpopulations. While recent studies of population statistics estimate that “50.4 percent of our nation’s population younger than age 1 were minorities as of July, 2011” (U.S. Census Bureau, 2012, pr.1), the majority of EL students live in poverty, come from families with lower levels of formal education, and struggle with a pattern of poor achievement of educational outcomes (Aud et al., 2011, 2012). For the U.S., the future well-being and educational prospects of EL children are of foremost concern because, although their well-being is important for the general health of the society (World Health Organization, 2015), the U.S. also needs an educated young population that is ready to meet the challenges of a rapidly changing, complex, global, and knowledge-based economy (Carnevale, Smith, & Strohl, 2010).

Over the last couple of decades, research has documented the poor performance of children who live in poverty relative to their peers from wealthier families with respect to

cognitive, social-emotional, and physical outcomes (Dearing, 2008; Reardon, 2011; Smeeding, 2016; Yoshikawa et al., 2012). This difference is most clearly observed in cognitive development. Children from poor families are more likely to fall behind in school readiness, to score lower on achievement tests, and to fail to graduate from high school or attend college (Brooks-Gunn & Duncan, 1997; Leventhal & Brooks-Gunn, 2000; Reardon, 2011; Sastry & Pebley, 2010).

EL students follow a similar pattern, as the results of the National Assessment of Educational Progress's (NAEP) long-term trend study for reading and mathematics indicates EL students do not fare well in comparison with their English-proficient peers (NAEP, 2012a, 2012b). As shown in Figures 1.1 and 1.2 respectively, over the last decade EL students in Grades 4, 8, and 12 have consistently scored statistically significantly lower than English-proficient peers in reading and mathematics (NAEP, 2012a, 2012b)¹. The achievement gap between the two groups increased at the last administration of the long-term NAEP assessment for all three grade levels both in reading and mathematics.

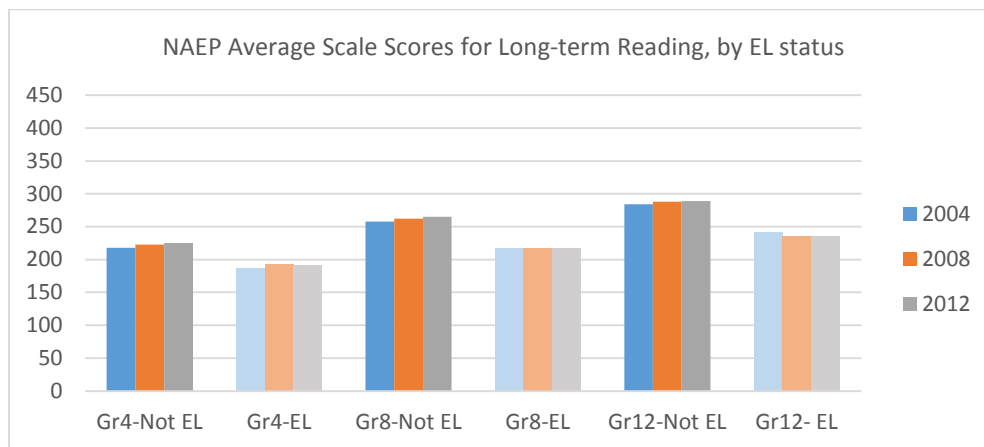


Figure 1-1. NAEP average scale scores for long-term reading, by EL status.

¹ The results from the main NAEP assessments suggest similar results. In 2004, 2008 and 2012 administrations, EL students in Grade 4, 8, and 12 have consistently performed statistically significantly lower than the not-EL students in reading and mathematics.

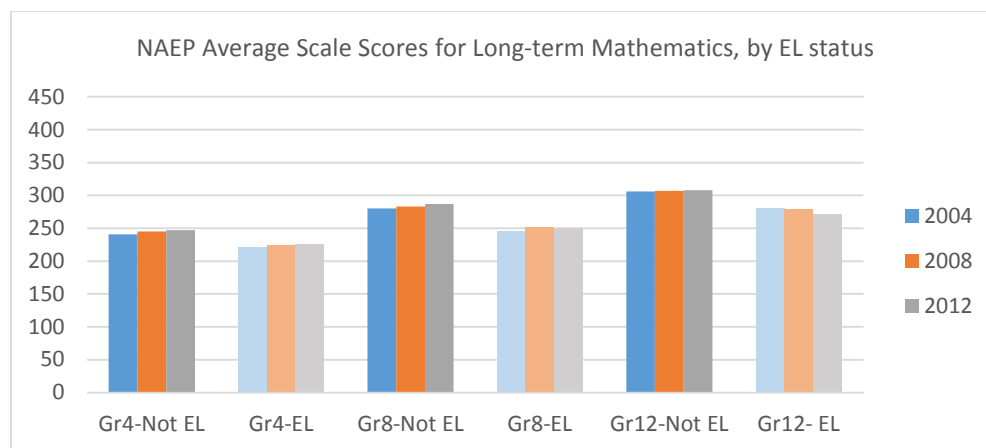


Figure 1-2. NAEP average scale scores for long-term mathematics, by EL status.

In contrast to their English-proficient peers, EL students who graduated from high school are also less likely to complete the core curricula in mathematics and science (Aud et al., 2012). Table 1.1 displays the percentages of EL and non-EL high school graduates who completed courses in specific STEM disciplines for the years 1990, 2000, 2005, and 2009. Although the completion rates of both English-proficient and EL students increased over these years, the differences between the two remained steady at about 10 percent for all four years (Aud et al., 2012). Also, little information on high school dropout rates for EL students is available, Aud et al. (2012) reported that high school dropout rates for students born outside of the U.S. (18.4 %) were nearly three times that of those born in the U.S. (6.5%) in 2010. In addition, data reported by EDFacts (2016) suggests that the public high school 4-year adjusted cohort graduation rate was 83.2% for the U.S. in school year 2014-15, while it was 65.1% for Limited English Proficient (LEP)² students in the U.S. A recent study of ELs in New York City, also found that

² Throughout this study, the author uses the term “English learners (ELs)” to refer to “Limited English Proficient (LEP)” students in public school systems. As August & Hakuta (1998) point out, I view the former as having a more positive connotation than the latter. However, because reclassification as “English proficient” constitutes a change in status, I will use the term “LEP status” to refer to this particular event in the life cycle of EL students in public school systems.

nearly 64 percent of all students who entered New York City public schools in grade 5 or 6 in the 2003-04 school year as EL graduated on time (i.e., earning any type of diploma within four years of entering grade 9 for the first time), which was seven percent lower than all students in the New York City public schools (Kieffer & Parker, 2017).

Table 1-1. *Percentage of high school graduates who completed specific STEM courses by year and EL status.*

	1990		2000		2005		2009	
	EL	Not EL	EL	Not EL	EL	Not EL	EL	Not EL
Algebra	66.6	77	62.3	66.5	63.7	68.4	68.8	73.3
Geometry	42.2	64.2	57.8	78.5	70.1	84.3	76.2	88.5
Algebra II/trigonometry	37.1	53.7	45.8	68.6	48	72	58.1	76.2
Analysis/precalculus	‡	13.4	15	26.7	13.8	30	19.4	35.6
Statistics/probability	‡	1	‡	5.7	3.9	7.9	4.4	10.9
Calculus	‡	6.6	‡	11.6	6.1	13.8	4.7	16.2
Biology	70.5	91.4	73.4	91.3	81.4	92.9	86.9	95.7
Chemistry	‡	49.3	34.9	62.1	46.1	67	47.4	70.9
Physics	‡	21.3	20.8	31.4	20.2	33.2	23.2	36.3
Biology and chemistry	‡	48	31.3	59.5	43	64.9	43.8	68.7
Biology, chemistry, and physics	‡	18.8	11.2	25.2	13.8	27.6	15.4	30.3

‡ Reporting standards not met. Either there are too few cases or the coefficient of variation (CV) is 50 percent or greater.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Transcript Study.

EL students also face the ongoing challenge of learning a second language. Research suggests that the process of acquiring a second language is different for each child. While a student typically requires between four and seven years to acquire and be capable of efficiently using academic language (Cummins, 1979; Hakuta, Butler, & Witt, 2000), acquisition of social language requires only about three years (Cummins, 1979). Sadly, both general education teachers and English as second language (ESL) teachers are not adequately prepared to teach EL students. On the one hand, general education teachers often lack the knowledge and skills to meet the linguistic needs of students from culturally and linguistically diverse backgrounds

(Artiles & Ortiz, 2002; Kushner & Ortiz, 2000; Zehler et al., 2003) and, on the other, ESL teachers typically lack content knowledge and instructional skills to effectively teach core courses in math and English language arts (Gersten & Baker, 2000).

In addition, research on college and career readiness suggests that students must complete a rigorous high school core curriculum in reading and mathematics to succeed in high school and beyond (Achieve Inc., 2004; Conley, 2007; Hein, Smerdon, & Sambolt, 2013). However, EL students are more likely to attend inferior schools with high student-teacher ratios, higher levels of students living in poverty, and low graduation rates and achievement levels in standardized assessments (Fry, 2008; Orfield & Lee, 2006; Sánchez, Ehrlich, Midouhas, & O'Dwyer, 2009) (Sánchez et al., 2009). Also, in the past five years, public high schools in the U.S. have been criticized for failing students to prepare for college and career. This is evidenced by the following: 1) high rates of remedial courses taken by first-year undergraduates enrolled in two- and four-year postsecondary institutions (Aud et al., 2011; McCabe, 2000; Sparks & Malkus, 2013), 2) employer dissatisfaction regarding high school graduates' deficiencies in the areas of basic and applied skills³ for entry level jobs (The Conference Board, 2006), and 3) a discouragingly low rate of on-time graduation from postsecondary institutions (Aud et al., 2013). Thus, schools pose additional challenges, increasing the difficulty for EL students to perform well throughout their schooling experience (Alliance for Excellent Education, 2011).

Given the abundance of research on the multiple challenges that EL students face and their higher-risk of academic failure, this study will examine whether receiving systematic non-

³ Basic knowledge/skills include English language (spoken), reading comprehension (in English), English writing skills (grammar, spelling, etc.), mathematics, science, government/economics, humanities/arts, foreign languages, and history/geography. Applied skills include critical thinking/problem solving, oral communications, written communications, teamwork/collaboration, diversity, information technology application, leadership, creativity/innovation, lifelong learning/self-direction, professionalism/work ethic, ethics/social responsibility.

academic student support during elementary school through the *City Connects* intervention affects EL students' likelihood of exiting LEP status during the elementary grades.

1.1 *City Connects* Intervention

Researchers at Boston College developed the intervention *City Connects* in response to two overriding concerns. First, in prior research, out-of-school risk factors—such as those associated with poverty as homelessness, violence, inequality in and lack of access to institutional resources—accounted for two-thirds of the variation observed in data on student achievement with teacher and school effects accounting for the rest (Berliner, 2013; Rothstein, 2010). Second is an inadequate student support system which 1) typically focuses only on students who are struggling either academically or behaviorally (Logan et al., 2015), 2) is usually limited in types of services provided, with a few connections to community partners (Walsh & DePaul, 2001), and 3) lacks standardized service-delivery practices across schools even within the same district (Lean & Colucci, 2010).

Recognizing the need for improved student support systems in urban schools located in high-poverty areas to address the myriad of out-of-school barriers to learning, the *City Connects* intervention was first implemented in Boston Public Schools (BPS) in 1999 (City Connects, 2014; Walsh et al., 2014). It targets out-of-school barriers to learning through prevention, intervention, and enrichment programs delivered through a network of school, family, community, and university partnerships. *City Connects* is based on a theory of change that regards out-of-school factors as barriers to a child's thriving in both cognitive and non-cognitive domains associated with school. Once provided with high quality resources and individualized services to meet their needs and foster their strengths, children will then, it is hypothesized, be cognitively, socio-emotionally, and physically ready to thrive in school. Thus, *City Connects*

hopes, through the significant improvements brought about through its interventions, to see improved academic achievement.

An empirical study examined the impact of *City Connects* intervention on students' report card scores and standardized achievement scores on the Massachusetts Comprehensive Assessment System (MCAS) English Language Arts (ELA) and Mathematics tests (Walsh et al., 2014). Although most findings related to Grade 3, 4, and 5 improvements on report cards and MCAS test scores were not statistically significant, treatment effects were, in general, positive. Walsh et al. (2014) found a significant school-level treatment effect on Grade 5 mathematics report card scores for *City Connects* students, and subsequent analyses of middle school data suggested lasting impacts even after *City Connects* students left the intervention (Walsh et al., 2014). With respect to Grades 6 and 7, significant and positive treatment effects at the school level were observed for MCAS Mathematics and Grade 6 MCAS ELA scores. In all the analyses, students enrolled in non-*City Connects* schools within the BPS system comprised the control group.

A dissertation study was also conducted on the treatment effects associated with *City Connects*. Employing a quasi-experimental design, Lee-St. John (2012) estimated the causal effects of the *City Connects* treatment on the likelihood of students being retained in Grades 6-8. *City Connects* students' overall probability of being retained in these grades was half (or 3.4 percentage points lower than) that of comparison students.

Recently, Dearing et al. (2016) studied the effects of *City Connects* on math and reading achievement of first generation immigrant students. This study followed multiple cohorts of students longitudinally and included schools which implemented the intervention at the beginning of school years 2001, 2002 and 2007. This quasi-experimental study revealed that

students who attended intervention schools with student support services achieved better Grade 5 math and reading test scores than students who attended comparison schools, after adjusting for student characteristics and early achievement. Finally, Dearing et al. (2016), examining the achievement gap between EL immigrant students who were proficient in English and those who were not, found this gap to be statistically non-significant for immigrant ELs who attended intervention schools.

1.2 Purpose of Study and Research Questions

The purpose of this study is to examine the treatment effects of the *City Connects* intervention on EL students' likelihood of exiting LEP status. Research suggests that exiting LEP status is an important educational indicator that often translates into improved educational opportunities (Abedi, 2008a; Francis & Rivera, 2007). One relevant area of study relates to college and career readiness. Research suggests that completion of, and high performance levels in, a rigorous high school core curriculum emphasizing reading and mathematics are strong predictors of college and career readiness (Achieve Inc., 2004; Conley, 2007; Hein et al., 2013). However, research has also established that EL students still classified as LEP are placed in classes that are less demanding with respect to academic content unless they can demonstrate that they are English proficient and can successfully function in mainstream classrooms (Garcia, 1999; Parrish et al., 2006). These findings imply that in order for EL students to have improved educational opportunities, they need to achieve English language proficiency and exit LEP classification prior to high school.

Additionally, research on time to reclassification into mainstream classes suggests that it usually takes between four and seven years to exit LEP status (Abedi, 2008a; Cummins, 1981; Grissom, 2004; Hakuta et al., 2000; Mavrogordato, 2012; Parrish et al., 2006; Slama, 2012;

Thompson, 2012). However, it is also argued that EL students should exit LEP status at the right time based on the correct evidence documenting their eligibility because of the academic consequences of early and late reclassifications as well as of misclassification into special education programs. For example, it has been documented that some EL students exit LEP status earlier than they should as the personnel involved with the reclassification decision confuses students' proficiency in social language with academic language, which takes a longer time to develop (Linguanti, 2001; Parrish et al., 2006). Furthermore, some EL students are misclassified into special education programs because of lack of appropriate tools to discern between a student struggling learning a second language and a student with learning disabilities (Sánchez, Parker, Akbayin, & McTigue, 2010). Furthermore, some EL students remain in the language support programs longer than the average time suggested by the literature (i.e., more than seven years). The causes of late reclassification is not clear as to whether the language programs that ELs attend are inferior (Flores et al., 2009) or ELs remain in these programs so long that they lose valuable time learning the grade-level academic content in mainstream classrooms (Mahoney & MacSwan, 2005).

This study hypothesizes that EL students in *City Connects* schools should exit LEP status sooner and at greater rates than their counterparts in the comparison schools, and thus, may have improved chances to complete the core curricula in high school for two reasons. First, given the characteristics of the student support system *City Connects* puts into practice in schools, one reason is that teachers might be more aware of their students' strengths and needs due to individual student and whole class review processes in the *City Connects* schools. In other words, teachers might be better able to monitor EL students over time and collect better evidence documenting their eligibility for reclassification.

The second reason is that EL students comprise a student subpopulation characterized as most at-risk to fail academically since its members face multiple challenges and are vulnerable to the risks of poverty (Kominski, Jamieson, & Martinez, 2011; Sheng, Sheng, & Anderson, 2011). Because the *City Connects* intervention is designed to meet the needs of students fitting the profile of most EL students (i.e., students who live in poverty), these EL students can be hypothesized to be highly likely to benefit from this intervention. The *City Connects* meet the needs of each student by providing them with a set of tailored prevention, intervention, and enrichment services through a network of school, family, community, and university partnerships. As these support programs are designed to address out-of-school barriers to learning, EL students in *City Connects* schools are expected to demonstrate improved readiness to thrive in classes resulting in both improved academic and non-academic success. As a result, it is important to conduct rigorous and scientifically based research to gain a better understanding of the extent to which the *City Connects* intervention is indeed associated with the mainstreaming of EL students. This study will investigate this association with the following research questions:

1. At each grade level, what proportion of students exit LEP status before the next grade in *City Connects* schools and in comparison schools?
2. To what extent is the *City Connects* intervention associated with students' likelihood of exiting LEP status while in elementary school after adjusting for student characteristics?
3. To what extent do *City Connects* and non-*City Connects* students differ in the median time needed to exit LEP status?
4. How robust are the estimated treatment effects to the presence of unobserved selection bias?

1.3 Significance of the Study

The results from this study should be of interest to researchers, educators, and policymakers who are concerned with students from low-income families and thus are at a higher risk for academic failure. Empirical studies on mitigating the effects of poverty are scarce, particularly when the population of interest consists of EL students (Devaney, Ellwood, & Love, 1997). First, the results from this study will provide empirical evidence on whether mitigating out-of-school barriers to learning affect EL students' likelihood of exiting LEP status.

Second, students in school typically receive support only if they are struggling academically or behaviorally (Sánchez et al., 2010). Students not struggling in these ways are often overlooked, with little attention given to their needs and strengths. However, in the *City Connects* intervention, every child is evaluated and receives a set of tailored prevention, intervention, and enrichment services. Hence, the results from this study will be applicable not only to EL students that are apparently struggling, but also to all EL students.

Third, *City Connects* is an early-life intervention which begins as early as kindergarten and is then implemented throughout the elementary grades. The results from this study will provide evidence about whether this early intervention improves students' likelihood of exiting LEP status at a younger age and thus translates into their spending more time in mainstream classrooms. In other words, the results from this study will provide evidence about whether *City Connects* reduces the number of years spent in LEP status.

Finally, although scientific research on exiting LEP status for EL students is still growing (Abedi, 2008a), more studies are needed to provide empirical evidence as to student and school characteristics that are significantly associated with this reclassification and to estimate the

median time required to exit LEP status. Thus, findings from this study will advance the research on this topic.

1.4 Outline of the Dissertation

This dissertation comprises five chapters. Chapter 1 begins with a brief review of the literature on poverty and its effects on child development with a focus on ELs as the fastest growing student subpopulation. This section builds the argument that out-of-school barriers to learning, which are often directly linked to poverty, are often overlooked in studies of achievement gaps, even though they clearly have significant effects on children's academic success. Here, the *City Connects* treatment, which is designed to offset out-of-school barriers to learning, is introduced. This chapter continues with a delineation of the problem this study addresses, the research questions, and the significance of the study's potential findings within the larger educational context.

Chapter 2 reviews the relevant literature in three domains: 1) typical processes involved in identification, placement, and reclassification of EL students, 2) challenges in the reclassification of EL students as English proficient, and 3) the median time required for EL students to exit LEP status and how these may be related to contextual factors.

Chapter 3 describes the methodology employed to address the research questions of this study. This chapter begins with a discussion of the research design and plausible threats to internal validity. It then describes data sources, sampling strategies, the outcome of interest, and the variables used during the analysis. In the last section, description of analytic approaches that will be employed are presented in detail, such as estimation of propensity score weights and modeling strategies for discrete-time event history analysis. A sub-section is also devoted for a

detailed description of discrete-time event history analysis as it is the main method of analysis in this study.

Chapter 4 presents results from the analyses outlined in Chapter 3. The first section reports results from descriptive analyses for the three samples, the big analytic sample, sub-sample 1, and sub-sample 2. In the next four sections, results from each analysis presented to answer the research questions of the study. These include the results from life-table and Kaplan-Meier analyses, baseline equivalence with ATT weights, and one-level and two-level discrete-time event history models. The results from the models are followed by fitted hazard probabilities and survival probabilities. Finally, the last section reports the results from the sensitivity analysis using the final model generated by the two-level analysis.

Finally, Chapter 5 presents a discussion of the results and their implications within the context of EL students and *City Connects* intervention. This chapter concludes the dissertation with study's limitations, policy implications and recommendations for future research in the field.

Chapter 2. Literature Review

2.1 Typical Processes for Identification, Placement, and Reclassification of EL Students

EL students have constituted the most rapidly growing sub-population in U.S. public schools (U.S. Census Bureau, 2012). At different times, this rapid growth rate has stimulated efforts to improve federal laws so as to protect EL students' rights to an equal education. This section briefly reviews these federal laws and describes how states translated them into policies and practices concerned with EL students.

In 1964, the Educational Opportunities Section (Title VI) of the Civil Rights Act was enacted to protect students from any form of discrimination by public education institutions based on race, color, gender, national origin, religion, and disability (U.S. Department of Justice, 1964). In 1970, a review of school districts with large national-origin minority groups revealed violations of Title VI of the 1964 Civil Rights Act, identifying four common ways by which districts denied such equal educational opportunities to EL students (U.S. Department of Education, 2005). Thus, a memorandum was published on Title VI to clarify these four issues: 1) public schools must ensure meaningful participation of EL students in educational programs by providing educational opportunities that would help these students gain proficiency in English, 2) students' lack of English proficiency cannot be employed as the reason for classification into special education programs and rejection into college preparatory programs, 3) language acquisition programs should be effective such that these students become English-proficient as soon as possible and such that their enrollment in such support programs is not permanent, and 4) parents of EL students must be notified of any such school activities and, if necessary, the notifications must be provided in a language other than English (U.S. Department of Education, 2005).

In 1974, in a civil right case involving Chinese-American students attending schools in San Francisco, California, it was argued that San Francisco public schools were failing to provide language programs to facilitate acquisition of English language and, thus, EL students were being discriminated against on the basis of their national origin (Lau v. Nichols, 1974). Because of this failure, the San Francisco school system was viewed as denying these students the right to meaningfully participate in educational programs and, thus, as having violated the 1970 memorandum. This case “clarified that equality of opportunity does not necessarily mean the same education for every student, but rather the same opportunity to receive an education” (NCELA, 2006, pr. 4). After the U.S. Supreme Court ruled in favor of these EL students, the U.S. Congress passed the Educational Opportunities Act to clarify the practice of providing an equal education (*Educational Opportunities Act*, 1974). This act prohibits any state from denying equal educational opportunity to any individual “by the failure by an educational agency to take appropriate action to overcome language barriers that impede equal participation by students in an instructional program” (*Educational Opportunities Act*, 1974, Section 1703(f)).

To comply with these federal laws and to protect the rights of EL students, all states then formulated their own laws, policies, and practices. In general, however, policies and practices of differing states were very similar. Thus, EL students typically progress through three phases during their time in public schools: 1) identification as LEP students, 2) placement into appropriate language-acquisition programs, and 3) reclassification as English proficient into mainstream classrooms. The sub-practices included in these three phases drive many aspects of EL students’ schooling experiences.

To ensure equal educational opportunities, the first step involves identifying the primary language of each newly enrolled student. The No Child Left Behind (NCLB) Act of 2001 (Title IX #25) defines EL students as follows:

- (a) age 3 through 21
- (b) enrolled or preparing to enroll in an elementary or secondary school
- (c) not born in the United States or whose native language is not English
- (d) is a Native American, Alaskan Native, or a native resident of the outlying areas
- (e) comes from an environment where a language other than English has had a significant impact on an individual's level of English language proficiency
- (f) is migratory and comes from an environment where English is not the dominant language
- (g) has difficulties in speaking, reading, writing, or understanding the English language that may be sufficient to deny the individual the ability to meet the state's proficient level of achievement and the ability to successfully achieve in classrooms where the language of instruction is English, or the opportunity to participate fully in society (No Child Left Behind (NCLB) Act of 2001 (Title IX #25)).

Based on these guidelines, state educational agencies use a parent-completed home language survey to obtain information on students' language backgrounds (Abedi, 2008a). If the home language survey indicates that the student has a language background other than English, the second step is to administer an assessment of English language proficiency (ELP) to determine the student's level of English proficiency.

School districts in Massachusetts follow a similar two-step practice (DESE, 2013a). After administering a home language survey to determine EL students' primary languages, they then employ an Assessing Comprehension and Communication in English State-to-State (ACCESS) assessment to determine students' English proficiency (DESE, 2013a). The ACCESS assessment was developed by the World-Class Instructional Design and Assessment (WIDA) consortium, of which 35 other states currently are members and so, along with Massachusetts, use ACCESS (WIDA, 2015). Massachusetts has employed ACCESS since 2012; it previously used the Massachusetts English Proficiency Assessment (DESE, 2013a).

If a student is found to be entitled to English language development and support services, this student's status is then classified as LEP. The NCLB uses this term to explicitly identify the subgroup to which such students belong (*No Child Left Behind Act*, 2001), thereby facilitating educational agencies' ability to monitor achievement gaps between student subgroups and thus provide a measure of school accountability with regard to closing achievement gaps (*No Child Left Behind Act*, 2001). Federal laws require that EL students are provided with opportunities to meet the same academic standards as non-EL students (*Educational Opportunities Act*, 1974, *No Child Left Behind Act*, 2001). Thus, states are required to develop programs based on established educational theory to provide these students with the opportunity of gaining English language proficiency, which would then facilitate their access to the regular curriculum and their opportunity to meet the same academic standards as non-EL students (U.S. Department of Justice, 1964). Massachusetts school districts provide one of four types of instruction for EL students: 1) sheltered English immersion, 2) English as a second language, 3) two-way bilingual education, and 4) transitional bilingual education (DESE, 2013a). At the third step, EL students are placed in one of these programs, and their parents are notified about their placement. Once notified, parents have the right to opt their children out of these language support programs (DESE, 2013a), and, should they avail themselves of this right, their children no longer receive instruction to support language development in English. However, opt-out students are still classified as LEP, have to participate in ACCESS assessments on an annual basis, and are monitored for reclassification (DESE, 2013a). Also, educational agencies are still required to ensure that these students have the opportunity to meet the same academic standards as non-EL students. Teachers are informed of students' placement decisions so that they can provide

sheltered English instruction or additional content area instruction through reading and math specialists for those who opted out but nonetheless require language support (DESE, 2013a).

In addition to providing all students with the opportunity to meet the same academic standards, federal laws require that EL students gain language skills as rapidly as possible and that participation in the support programs enabling them to do so be only temporary (U.S. Department of Education, 2005). At the state level, this policy translates into annual assessments for English language proficiency and decision making for placement and reclassification. In Massachusetts, districts are advised to form school-based teams to make placement and reclassification decisions for EL students (DESE, 2013a). These teams typically consist of students' teachers (both content area and English as second language), school personnel (school guidance counselors, psychologists, or special education teachers [if applicable]), and one administrator (assistant principal or principal) (DESE, 2013a). This team reviews existing data on a student, including information about the student's first language, number of years in U.S. schools, language assessment results, content area assessment results, English language proficiency level, grade level, special education status (if applicable), teacher observation notes, grade progress reports, and any other information related to the student's general performance in school (DESE, 2013a). For the reclassification decision, EL students must demonstrate proficiency in ACCESS and, if available, in MCAS ELA (DESE, 2013a). Results from these assessments are considered an indication of students' ability to perform ordinary classwork in English (DESE, 2013a). However, school-based teams consider the results from these assessments in conjunction with other relevant student data to render the final decision on reclassification (DESE, 2013a). In other words, in some cases, students may not be reclassified

into mainstream classrooms even though they have demonstrated proficiency both on ACCESS and MCAS ELA.

2.2 Challenges to Reclassifying EL Students into Mainstream Classrooms

EL students have been the subject of research for the last four decades (Genesee, 2006). However, the research is equivocal on what actually constitutes being English proficient and on the criteria that EL students should meet to be reclassified as English proficient (August & Shanahan, 2006; Kindler, 2002; Mahoney & MacSwan, 2005; NCELA, 2007). This ambiguity is also present in states' reclassification policies and practices (Linguanti, 2001). States only provide general guidelines for a reclassification decision, with the exception that students have to perform at certain proficiency levels in ELP assessments and state ELA standardized achievement tests. Reclassification decisions are made by the school-based teams on a case-by-case basis once these students meet the benchmarks associated with required assessments. Ultimately, this leads to different conceptions of "being English proficient" and depends on the type and quality of evidence used to demonstrate students' English language proficiency. The following paragraphs outline the problematic aspects of the reclassification process, which have been commonly highlighted by the relevant research.

2.2.1 Quality of ELP assessments has implications for reclassification.

Since states use ELP assessments to identify, monitor, and reclassify EL students, the validity and reliability of these assessments are of foremost concern (Abedi, 2008b). The literature distinguishes among ELP assessments based on when they were developed, either pre-NCLB or post-NCLB. In the following, the concerns with respect to post-NCLB ELP assessments are discussed since NCLB was in effect for the years studied in this dissertation.

With the NCLB Title III, four multi-state consortia emerged to construct the new generation of ELP assessments. The new assessments included updated ELP content standards and they were designed to measure English proficiency in four domains: reading, writing, speaking and listening. Also, where possible, ELP standards were aligned with the content area standards in major academic topics, such as English, Math, Science and Social Studies (*No Child Left Behind Act*, 2001)

Although all these were positive improvements, the new generation of ELP assessments was still criticized in the literature on multiple grounds. Firstly, as of 2006, 25 different ELP assessments were being used by states (NCELA, 2007). Different tests may lead to different results and, hence, different academic consequences (Abedi, 2008b). For example, Fast, Ferrara, and Conrad (2004) reported that states followed different standard-setting approaches during the development of the new generation of ELP assessments, and Abedi (2008b) reported that such differences “may lead to different interpretations of students’ level of ELP” (pg. 198). As a consequence, a student who performed at the intermediate level in one state could be categorized as proficient in another state. Summarizing additional concerns relating to post-NCLB ELP assessments, Abedi (2008b) listed 1) English language content standards that are not clearly defined in all states; 2) reporting of composite scores from the four subscales (i.e. reading, writing, speaking, and listening) when a student performs very poorly in one area but adequately in other areas; and 3) confusion as to whether these tests should measure English language that facilitates content learning or academic content itself.

Some of the above concerns were alleviated by the development of ACCESS assessment by the WIDA consortium, of which Massachusetts and 34 other states currently are members (WIDA, 2015). Massachusetts has employed ACCESS since 2012; it previously used the

Massachusetts English Proficiency Assessment (DESE, 2013a). The initiative to design ACCESS began in 2002 and the test became operational for the first time in 2005 in three states. It was then rapidly adopted by other states over the years. A description of the ACCESS outlining its characteristics was provided by Bauman, Boals, Cranley, Gottlieb, & Kenyon (2007, pg 83) :

- Anchored in WIDA’s English language proficiency standards
- Aligned with core academic content standards
- Vertically scaled across grade level clusters
- Divided into tiers within each grade level cluster to accommodate a range of contiguous proficiency levels
- Includes listening and speaking domains in addition to reading and writing

Massachusetts Department of Elementary and Secondary Education states that the WIDA English Language Development Standards are an important component of the Department’s Rethinking Equity and Teaching for English Language Learners (RETELL) initiative, which is designed to strengthen the teaching and learning of ELs and address proficiency gaps (DESE, 2013b). The following is provided by the Department explaining the reason for adopting WIDA standards and ACCESS:

[T]hey provide useful data and research-based resources for promoting language development along content area learning. In addition, they provide a common language between content, vocational and language teachers to maximize collaboration on behalf of ELLs (DESE, 2013b. pg. 3).

2.2.2 Standardized achievement tests not designed for reclassification purposes.

Reclassification decisions depend not only on English language proficiency but also on academic achievement (Kindler, 2002). Students’ performances on statewide standardized academic assessments are viewed as an indicator of success in English-only classrooms. For this reason, state guidelines often require EL students to meet certain proficiency levels in statewide academic assessments.

However, for several reasons the use of state standardized assessments in the reclassification decision is problematic. The first is the use of different proficiency level criteria by different school districts within the same state. Several research studies have reported that different proficiency levels were used for reclassification decisions within the same state, suggesting that a student considered proficient in one school district may not be considered proficient in another district (Gandara, 2000; Grissom, 2004; Linqunti, 2001). For example, in Massachusetts, while the State Department of Education guidelines state that school-based teams must consider the student's performance on MCAS content area tests, it does not suggest a minimum benchmark. The following is provided for the school-based teams to make a more informed decision:

[U]nless an ELL student did not participate in MCAS ELA testing either because he or she is a student in kindergarten through grade 2, or is a first-year ELL student and was not required to participate, or participated instead in the MCAS-Alt, the most recent MCAS ELA results should serve as a key indicator of the student's likelihood of performing ordinary class work in English. Those results should be used to support and validate the preliminary decisions made each spring about the student's instructional programming and ELL classification (Chester, 2013, pg.12).

The second major concern raised was that these tests are designed to measure the content knowledge of native English-speaking students and not EL students (Rossell, 2000; Stefanakis, 1998). Abedi (2006) reported increased levels of construct-irrelevant variance for LEP students compared to native English speakers on standardized assessments due to the linguistic complexity of the test items. In other words, students' low performances on these tests were not necessarily due to lack of knowledge but sometimes to the construct-irrelevant linguistic complexity of the items. Construct-irrelevant variance undermines the validity of the intended inferences from these assessments as it increases measurement error (Haladyna & Downing, 2004; Messick, 1994). Third, Abedi (2008a) questioned the use of these tests from the perspective of native English speakers. The definition of LEP reclassification becomes

problematic when many native English speakers also score low on these standardized achievement tests. When native English-speaking students score low on standardized tests it is assumed that they lack the content knowledge measured by the test. However, construct-irrelevant linguistic complexity might also contribute to their low scores. Thus, it becomes problematic to categorize EL students as not English-proficient when both EL and native English-speaking students might be suffering from the same issues related to the test.

2.2.3 Lack of a widely accepted second language acquisition theory.

Another aspect of the criticism of the reclassification process is the absence of a widely accepted second language acquisition theory. Two main schools of thought exist with regard to second language acquisition (Conteh-Morgan, 2002). One school highlights the technical aspects of learning a new language and emphasizes individuals' growing awareness of the new language's grammar and vocabulary from the perspective of their native language (Chomsky, 1968; Krashen, 1988). The other school emphasizes psychological aspects and so views environmental factors as playing an important role in the acquisition of the new language. Environmental factors include exposure to rich learning environments and social interactions with native speakers (Reutzler & Cooter, 2007).

Without a generally accepted scientific explanation of the process whereby a second language is learned, understanding the instructional needs of students who are LEP is difficult. Also, school-based teams are unclear as to the criteria against which they should interpret relevant student data. For example, two studies are often cited in the literature regarding the amount of time ELs require to gain proficiency in basic interpersonal communicative skills (BICS) and cognitive academic language proficiency (CALP), respectively. These studies suggest that, while about three years are required to become proficient in BICS, approximately

four to seven years are needed to gain proficiency in CALP (Cummins, 1979; Hakuta et al., 2000). Although empirical data on the estimated time required to become proficient in a second language is limited, these studies have led to many discussions on the academic consequences of early and late reclassifications of EL students as well as of misclassification into special education programs.

With the distinction between BICS and CALP clarified, it became evident that EL students' proficiency in BICS was oftentimes confused with proficiency in CALP, resulting in early reclassification of ELs. Students subject to this early reclassification were thus provided with less language support services in subsequent grades, thereby increasing their likelihood of academic failure (Linguanti, 2001; Parrish et al., 2006). Studies suggest that EL students reclassified during elementary grades struggle to close the academic achievement gap with their native English-speaking peers in upper grades (Flores, Painter, Harlow-Nash, & Pachon, 2009; Gandara, 2000; Jong, 2004).

On the other hand, some students were identified as remaining in the language support programs longer than the average time suggested by the literature (i.e., more than seven years). Such students are often referred to as "long-term ELs" (New York City Board of Education & Accountability and Assessment, 2000), and there is an ongoing debate as to the causes of their academic struggle: Are the language programs they attend inferior (Flores et al., 2009), or do they remain in these programs so long that they lose valuable time learning the grade-level academic content in mainstream classrooms (Mahoney & MacSwan, 2005)? For example, in California, 70% of EL students were in LEP status for more than five years, and their likelihood of reclassification decreased after that time (Grissom, 2004; Parrish et al., 2006). In New York City, about 10% of LEP students were characterized as long-term LEPs and retained this status

for seven or more years (New York City Board of Education & Accountability and Assessment, 2000).

Finally, Sánchez, Parker, Akbayin, and McTigue (2010) found that school-based teams had difficulty in distinguishing between a student who was LEP and struggling to acquire a second language and one who had learning disabilities. The difficulty in making this distinction resulted in a disproportionate representation of EL students in special education (Garcia & Ortiz, 1988). A review by Sánchez et al. (2010) of districts' referral processes to special education programs for students who were LEP revealed four main challenges: 1) school-based teams had difficulty interpreting policy guidelines, 2) members of the school-based teams held differing views about the necessary time in LEP programs before a student could be considered for referral to special education programs, 3) members of the school-based teams had insufficient knowledge in both second-language acquisition and in identification of learning disabilities, and 4) access to assessments that could reliably differentiate between second-language development and learning disabilities was lacking.

2.2.4 Contradictory NCLB incentives for reclassification of EL students.

Reclassification of EL students into mainstream classrooms also has implications for schools' accountability measures with respect to NCLB's adequate yearly progress (AYP), since NCLB was in effect until before Every Student Succeeds Act (ESSA) passed in December 2015. NCLB defines EL students as an explicit subgroup under the name of "LEP students" (*No Child Left Behind Act*, 2001). However, NCLB Title I and Title III requirements create contradictory incentives for schools to reclassify LEP students. On the one hand, NCLB Title I mandates schools to report results from statewide standardized tests in ELA, mathematics, and science broken down by all student subgroups. This policy pressures schools to keep their top

performing EL students in LEP status in order to meet AYP requirements for this subgroup (Kieffer, Lesaux, & Snow, 2008). On the other hand, NCLB Title III rewards schools for high reclassification rates (Kieffer et al., 2008), thus encouraging schools to reclassify EL students as early as possible to show that a greater number of students meet the LEP proficiency goal. This becomes an issue for the EL subgroup only because, by design, this is the only subgroup whose composition changes based on reclassification. Linqunti (2001) describes this problem as a redesignation dilemma since schools are provided with conflicting incentives based on either late reclassification to inflate subgroup performance (Title I) or early reclassification to demonstrate efficacy of LEP programs (Title III). With the passage of Every Student Succeeds Act (ESSA) in December 2015, schools are still required to report achievement scores for ELs as an accountability measure. However, accountability measures concerning ELs are all moved under Title I, thus, eliminating the funding conflict (National Conference of State Legislatures, 2015).

2.3 Research on Reclassification of EL Students as English Proficient

Only a small number of studies have examined the average time needed for EL students to be reclassified as English proficient. This section provides a brief overview of these studies and then discusses major contributions and critical issues related to the reclassification event.

The research on the average time needed for students to become proficient in English can be traced back to late 1970s. In 1977, Oller coined the term global language proficiency, described as a one-dimensional construct accounting for majority of the variance observed in language proficiency. Cummins (1979) criticized Oller (1977), whose definition of language proficiency does not distinguish language skills required for daily communications from skills required for academic learning. Cummins (1979), thus, advocated a two-dimensional construct: one reflecting BICS, a term more commonly associated with ‘social’ language, and the other

reflecting CALP, applicable to academic language. To demonstrate the distinction between BICS and CALP, Cummins (1981) reanalyzed Ramsey and Wright's (1974) data from 1200 randomly selected EL students in Grades 5, 7, and 9 of the Toronto school system. Cummins (1981) examined the relationship between age on arrival in Canada, length of residence in Canada, and scores on language tests. The findings suggested that, regardless of age on arrival, after five years of residence in the host country, EL students began to approach grade-level norms for the tests that appear to measure CALP. After five years, Cummins (1981) reported, the effect of length of residence started to flatten around the grade mean. Cummins (1981) concluded that older and younger learners exhibit similar trends in their progression towards grade-level norms since analyses detected no significant relationship between age on arrival to the host country and amount of time before EL students were categorized as English proficient. Furthermore, Cummins (1981) reported a similar trend but with a different time frame for tests that measure BICS. After three years of residence in the host country, EL students became proficient in BICS, and then the scores began flattening around the grade-level norms. In other words, Cummins (1981) found that more than three years of residence did not have a significant effect on BICS scores.

Cummins (1981) is considered a pioneer in the study of the distinction between proficiency in BICS and CALP. Equally important, this was also the first study to examine the relationship between length of residence, age on arrival, and the time needed to be classified as English proficient. Thus, this study provided the initial scientific evidence in researchers' and educators' quest to understand when EL students are more likely to exit LEP status and can be reclassified into mainstream classrooms.

However, before making any generalizations to the greater EL population, the limitations of this research deserve examination. First, the author notes concerns with regard to the validity of the tests designed to measure BICS. Particularly, Cummins (1981, pg. 135) notes that the “English Competence Test was an experimental test developed by the Toronto Board,” and early analyses showed that some of its parts “were tapping only some aspects of performance.” Second, as noted by Cummins:

[F]indings are not necessarily generalizable outside the Canadian social context, and even within that context may not hold for particular immigrant groups. A complex array of social, educational, affective and cognitive factors determine second language acquisition by immigrant children and differences in these factors and their interactions will be reflected in differences in patterns of second language acquisition (Cummins, 1981, pg. 148)

Finally, student mobility is an important factor in estimating the length of time required for EL students to become English proficient, but Cummins's (1981) study does not account for student mobility. Thus, an analysis based on a sample that excludes mobile students may have resulted in underestimation of the length of time required to become English proficient.

The second study on the average time required for EL students to exit LEP status was conducted by Hakuta et al. (2000). This work was based on data collected from two school districts in the San Francisco Bay area. The two districts were similar with respect to number of EL students but differed in the percent of students who received free- or reduced-price lunches (35% District A, 74% District B), the type of English support programs offered (English as second language in District A and bilingual education in District B), and the predominantly spoken language (Vietnamese in District A and Spanish in District B). In both districts, the samples included only those students who had been enrolled since kindergarten and identified as LEP. In District A, 1872 students were included in the study, and in District B, 122 students were randomly selected from students who met the criteria.

In District A, results from three annual assessments were examined: the Idea Proficiency Test (IPT) for English, the MacMillan Informal Reading Inventory test, and a district-developed writing assessment. In District A, students' reclassification designations were also available for analysis, which were made by school teams based on the results from these assessments and relevant student data. In District B, only students' scores for the Woodcock Language Proficiency Battery (Revised), in Grades 1, 3, and 5 were available; these measured English proficiency in oral language, reading, and writing.

Analyzing these two samples separately, the findings of Hakuta et al. (2000) suggested that 90% of District A students scored proficient in the oral part of the IPT by the end of Grade 4, whereas for the reading and writing assessments, 90% of the students achieved proficiency between the end of Grade 4 and Grade 6. Also, by the end of 4th grade, more than 40% of students were reclassified as English proficient. In District B, where EL students' performances on the tests were compared to the norm for native English speakers, Hakuta et al. (2000) found that EL students were one year behind age-equivalent performances for Grades 1 and 3 but that this gap widened to two years in Grade 5. For both districts, these analyses were repeated by dividing the samples into categories based on socio-economic status (SES), and results indicated high positive correlations between low SES and later attainment of proficiency in English.

Hakuta et. al's work (2000) is unique; it not only provided empirical evidence on when EL students in the U.S. were more likely to score proficient in English assessments and be reclassified as English proficient, but, by including SES in the analysis, also was the first study to explore the relationship between SES and reclassification.

Although the findings from this study are in part consistent with those reported by Cummins (1981), they are subject to some methodological concerns. First, the authors do not

provide detailed information on the statistical methods used. Therefore, it is difficult to judge the rigor of their methods and the extent to which findings were statistically significant. Second, as Cummins (1981) pointed out, generalizability of the findings is problematic since context, represented by community, may differ significantly. Third, as Hakuta et al. (2000, pg. 8) mention, the caveat in this study concerns mobile students, which make the sample “more selective as the grades go higher, because students move away from the district.” In other words, the methods fail to account for mobile students and so, as in Cummins's (1981) study, the times to proficiency were likely to be underestimated.

In 1998, California passed Proposition 227, which dictated significant changes in the state's laws concerning the education of EL students (California Law, 1998). Among these changes were the requirement that “all public school instruction to be conducted in English” and that EL students be placed in English-acquisition programs only for a short term, “not normally exceeding one year” (California Law, 1998, pr.1). While these changes led to sheltered English-immersion programs becoming more widespread, two studies suggested that Proposition 227 was not successful for the rapid transitioning of EL students out of LEP status within one year (Grissom, 2004; Parrish et al., 2006).

Studying the impact of Proposition 227 on EL students' reclassification rates, Grissom (2004) critiqued its usefulness by examining longitudinal data on three cohorts of students who attended California public schools from Grades 2 through 5 from 1998 through 2000. The sample excluded students who were retained in a grade, who left before the end of fifth grade, and who enrolled in the California public school system after the second grade. Grissom (2004) carried out three types of analyses for three cohorts of students (cohort 1 n=192,023 cohort 2 n=224,425, and cohort 3 n=277,373): 1) for each cohort, the percent of EL students who were

reclassified between the second and fifth grades was calculated; 2) a logistic regression analysis was conducted where whether students were reclassified or not was regressed on the student achievement score on a statewide norm-referenced test, gender, SES status (as measured by free- or reduced-price lunches), and native language status (Spanish vs. neither English nor Spanish); and 3) students' academic scores on statewide tests were examined by EL students' language category, that is 'English Only,' 'Fluent English Proficient,' 'Reclassified Fluent English Proficient,' and 'English Learner' (Grissom, 2004).

The goal of the first analysis was to examine reclassification rates longitudinally by following the same group of students from Grades 2 through 5. Study findings suggested that percent of reclassification increased as EL students progressed through the grades: It was lowest in Grade 2 (ranging between 1.4% and 2.2%) and highest in Grade 5 (ranging between 29.7% and 32.3 %). Grissom (2004, pg.10) states that "after four or five years of schooling only 30 percent of EL students had been reclassified" despite the state law requiring LEP designation not to exceed one year. The second analysis, in which the relationship between the probability of being reclassified and students' demographic characteristics were examined using logistic regression, Grissom (2004) found that, while girls were more likely to be reclassified than boys, students whose primary language was Spanish were less likely to be reclassified than students whose primary language was other than Spanish, after accounting for achievement. The third analysis examined average achievement scores longitudinally for each cohort broken by students' LEP category. Similar trends were identified for each cohort. The English learner category had the lowest average scores across subsequent grades when compared to 'Reclassified Fluent English,' 'Fluent English,' and 'English Only' students. However, as noted by Grissom:

[T]he continuously low academic performance of English Learner students should not be interpreted to mean that English Learner students never improve or were failing to close

the gap between themselves and the other language categories. Each year the English Learner group represented those students who were left behind after the most academically able were reclassified as Reclassified Fluent English (Grissom, 2004, pg.23).

Based on these findings, Grissom (2004) criticized Proposition 227, concluding that the goal, which was to have EL students exit LEP status in one year after their first enrollment into the state's school system, was not met. The highest rates of reclassification, only around 30%, happened in the fifth grade. Although Grissom (2004) argued against Proposition 227 based on the fact that early reclassification appeared to reduce learning opportunities with respect to academic English, he also raised concerns in regard to low reclassification rates by the end of elementary school, which was around 30%.

The findings from Grissom's (2004) study made important contributions to the examination of several aspects of the reclassification event. 1) He employed data that were longitudinal in nature and originated from the entire state, not just from a particular school district. 2) He used simple methods to study the research questions. 3) He examined student background characteristics as predictors of reclassification using a logistic regression analysis.

Although all these were intended to provide a clearer understanding of when reclassification was more likely to happen and whether certain student characteristics were more strongly associated with reclassification, it is important to note that the study was subject to important methodological weaknesses. First, the sample consisted of the same group of students who remained in the California public school system from Grades 2 through 5 and who were never retained in grade during this time frame. Thus, the study excluded both mobile and retained EL students from its analyses. As pointed out previously, failure to account for mobility likely biases estimated reclassification rates downward at a given point in time, as we do not know when such students experienced the reclassification event. Additionally, focusing on a

specific sample—not retained and not mobile students—raises the concern that the sample is systematically different from the general population of EL students. Next, the use of percent reclassified at each grade could be considered as a simplified approach to estimate reclassification rates, where such statistical techniques as survival analysis would have been more appropriate. Finally, although the data was drawn from the whole state, the analyses failed to account for the multilevel structure of the data. Schools, in general, differ from each other with regard to the characteristics of their student populations, in particular their EL populations. As a result, Grissom's (2004) study was unable to explore how reclassification varied across schools within the state.

The California Department of Education evaluated the impact of Proposition 227 in its fifth year as part of a legislative mandate (Parrish et al., 2006). The report prepared by Parrish et al. (2006) was very comprehensive, examining several aspects of Proposition 227 including problems with its implementation, its impact on EL students' academic achievement, and its impact on re-designation of EL students as English-proficient. Although Parrish et al. (2006) and Grissom (2004) examined the impact of this proposition to address similar research questions, Parrish et al. (2006) used discrete event history analysis, which is a superior method for estimating the time required for EL students to be reclassified as English proficient. This method accounts for mobile students, whose reclassification status and time of reclassification, if applicable, were unknown (Singer & Willett, 1993).

Parrish et al. (2006) used a sample drawn from a statewide student-level dataset which tracked students over years. The first phase of their analysis explored the probability of reclassification over time with an unconditional model. The second phase controlled for ethnicity—Hispanic, Asians, and Whites—to examine the extent to which ethnicity was associated

with the likelihood of reclassification. In the third phase, the researchers repeated the same analysis on a subset of the sample to determine whether reclassification rates over time varied across six school districts.

In the first-phase results, only 2.5% of EL students were reclassified within the first year of attending California public schools, but this cumulative rate increased to 25% after five years and to 40% after 10 years. “These results [were] very close to those reported by Grissom (2004), who found that proportion of English Learners not re-designated after five years is about 70 percent” (Parrish et al., 2006, pg. III-33). The second phase of the analysis indicated the presence of differences in the probability of reclassification based on student ethnicity. EL students from Hispanic backgrounds were significantly less likely to be reclassified in comparison to EL students from Asian and White backgrounds. Specifically, Hispanic EL students’ likelihood of reclassification was 26% after six years whereas it was 45% and 50% for Asian and White EL students, respectively. Finally, the results from the third phase of analysis indicated that likelihood of reclassification varied greatly by school district; in one case, probability of reclassification was four times that of another district.

Apart from concluding that Proposition 227 did not meet its goal of having EL students reclassified as English-proficient in one year, the findings from this study have significance for the field. Since this study is the first to employ discrete event history analysis as the method of analysis in estimating the time required for EL students to leave LEP status, its estimates are less biased than the ones obtained previously. Also, it was the first study to examine variation in reclassification across different ethnicities and districts employing more advanced methodologies.

Nonetheless, the authors themselves raised several concerns relating to limitations of their study. First, Parrish et al. (2006) point out that their analyses did not account for all the factors that might have been associated with variation in reclassification rates, including differing reclassification policies employed by different districts, types of language programs, students' primary languages, and poverty. Second, although the sample was drawn from the whole state, the multilevel structure of the data was not taken into account and thus might have caused estimates to be biased due to variations based on district. Third, Parrish et al. (2006) noted that at the time when Proposition 227 was implemented, other reforms had also taken place, particularly one involving reduced class sizes. As a consequence, disentangling the effect of Proposition 227 from the effect of the reduction in class size on reclassification rates were not possible (Parrish et al., 2006).

In the U.S., early studies on the reclassification of EL students have been primarily conducted in California, one of the states having the highest proportion of EL students (U.S. Department of Education, 2014a). Although these studies made important contributions to the field, they all had some methodological shortcomings, specifically in regard to the exclusion of mobile students. The area (i.e., California) in which studies were conducted limited the study's applicability; differing policies, practices, and compositions of EL students in different states could produce very different results. The next set of studies to be discussed was performed more recently and used discrete event history analysis with several covariates. Moreover, they provide empirical evidence both from California and other states in the U.S.

Abedi (2008, pg. 17) points out that improper classification of EL students due to invalid assessments “may lead to inappropriate and inadequate instruction for EL students.” Building on this notion, Abedi (2008) argues that valid EL identification and reclassification systems are of

the utmost importance as the academic consequences of improper classification are severe. Besides the use of ELP and standardized achievement tests, Abedi (2008) posits other determinants of EL student reclassification outcomes such as gender, SES, ethnicity, and parent educational level. Thus, he conducted a discrete-time event history analysis on a group of nearly 24,000 students whom he followed for six years from Grades 7 through 12. The study found that the probability of reclassification over time did not differ by gender or by poverty status (free or reduced price (FRPL) versus full price). However, students whose prior reading scores were high were substantially more likely to be reclassified earlier than those students having low reading scores. Similarly, students of Caucasian and Asian ethnicity were more likely to be reclassified earlier. Abedi (2008, pg. 25) states, "It took almost ten semesters for Hispanic students to be reclassified from EL to fluent English Proficient, while it took half as much time for Asian and Caucasian students to be reclassified." Although Abedi (2008) discusses the potential reasons for inconsistent reclassification systems, he does not provide a specific explanation about why linguistic background might be one of the factors contributing the differences in reclassification.

Abedi (2008) advanced the research on EL reclassification by incorporating several student characteristics into the discrete event history analysis. In addition to accounting for mobile students, his analyses provided new information regarding possible causes of variation in reclassification rates. However, one element in the study's design may have caused it to underestimate average reclassification times. The study examined the reclassification event starting from the seventh grade and so counted the time that a student was classified as EL as of this grade. However, some students in the study may have been classified as EL prior to the seventh grade, thereby lengthening the actual time they spent before reclassification.

For her dissertation research, Thompson (2012) conducted a longitudinal study using data from the Los Angeles Unified School District and estimated the average time required for EL students to be reclassified as English proficient. The analytic sample included nearly 203,000 EL students who were enrolled in the district for the first time as kindergarteners between the school years 2001-02 through 2009-10. Conducting a discrete event history analysis on this data, Thompson's (2012) model included several student-level variables: gender, ethnicity, home language, special education status, free- or reduced-price lunch status, participation in bilingual program, initial English-proficiency level, proficiency in primary language at school entry, and parental level of education. Thompson (2012) also created group-centered, school variables by aggregating student level data within each school. These included the percent of EL students, percent of Spanish speakers, and percent of students receiving free- or reduced-price lunches. Thompson (2012) summarized her findings as follows:

[B]oys, native Spanish speakers, students with lower levels of initial English proficiency, students with lower levels of initial proficiency in their primary language, students who ever qualified for special education, students who qualify to receive free or reduced-price lunch, and students whose parents have lower levels of education all have lower probabilities of reclassification than their peers, controlling for the other factors (Thompson, 2012, pg 11).

Additionally, the author examined the interaction between time and whether students were ever in a bilingual program. Thompson (2012) states that the underlying reason for examining this interaction was based on the findings from prior research on bilingual programs, which suggested that students in bilingual programs are more likely to be reclassified in later years because in early years such programs are more focused on development in students' primary language. In line with this, she found that:

[T]he negative coefficient on the main effect for having ever been in a bilingual program suggests that students ever in bilingual programs are less likely than their peers

to be reclassified. However, the positive coefficient on the interaction between whether students were ever in a bilingual program and time suggests that students ever in bilingual programs become increasingly likely to be reclassified in later years (Thompson, 2012, pg 41).

Moreover, the author reported that approximately four to seven years were required for EL students to be reclassified as English proficient and that their likelihood of reclassification decreased after the sixth grade.

Thompson (2012) ran a second set of discrete event history analyses in which she used only annual student ELP test scores in reading, writing, and listening-speaking and ELA scores from the statewide standardized assessment. According to the results, 89.7% of EL students met the reclassification criteria for the listening-speaking section of the test, which was designed to be at intermediate level, by their third year (i.e., when they were in the second grade). Students typically begin taking the reading and writing parts of the test in the second grade. Thompson (2012) found that it took much longer for EL students to meet the proficiency level required for reclassification for the reading and writing parts of the test compared to its listening-reading section. Reporting a similar pattern for meeting the required proficiency level on the ELA, Thompson (2012, pg. 50) concluded the following:

[T]he point at which 60% of students have met the criteria – the time necessary for students to reach proficiency on literacy-based measures ranges from four to five years (four years for the CST ELA and CELDT Writing criteria and five years for the CELDT Reading and CELDT Overall criteria). This is calculated simply by noting the time point at which the survival complement for each criterion exceeds the .6 level.

Thompson's (2012) findings provided a more recent picture of reclassification in California. In comparison to previous studies, her analyses accounted for many more student level variables, thereby helping to better explain variation observed in reclassification rates. Yet, Thompson's findings were, in general, consistent with those from previous studies. For example,

in Thompson's study, time to reclassification was four to seven years as in Hakuta et al. (2000), Grissom (2004), and Parrish et al. (2006). Parallel to findings of Abedi (2008) and Parrish et al. (2006), EL students from Spanish backgrounds were more likely to be reclassified at a later grade than were EL students from Asian backgrounds. This study was also the first in this area to account for the school clustering effect with group-centered variables.

Mavrogordato (2012) also used discrete event history analysis to examine the rate at which EL students were reclassified into mainstream classrooms and the average time required for EL students to be reclassified. The analytic sample included 58,269 first-grade EL students attending Texas⁴ public schools in the school year 2002-03. The study followed this cohort of students until the end of fifth grade, and the analysis included three types of variables: student demographic characteristics (gender, socio-economic status, native language), student educational profile (type of English support program, special education status, gifted status, number of school switches, whether retained in a grade in a previous year, and number of disciplinary infractions), and student achievement results in ELP and statewide standardized assessments.

Mavrogordato (2012, pg. 133) reported that students' achievement results were "by far the most powerful predictors of reclassification." Specifically, Mavrogordato (2012) found that students who met the ELA proficiency level in Texas statewide assessments were twice as likely to be reclassified in a given year. Additionally, in Texas, where bilingual programs are offered,

⁴ In Texas, "at the end of the school year, a district may transfer (exit, reclassify, transition) an English language learner (ELL) out of a bilingual or ESL education program for the first time or a subsequent time if the student is able to participate equally in a regular all-English instruction program as determined by satisfactory performance in all three assessment areas (i.e. Oral, Listening & Speaking Assessment, English Reading, and English Writing on State of Texas Assessments of Academic Readiness (STAAR) and the results of a subjective teacher evaluation" (Texas Education Agency, 2016)

EL students also have the option to take the achievement tests of reading and writing in their native languages. The purpose of this practice is to capture the academic performance in that content area as opposed to their English proficiency. Mavrogordato (2012) found that the probability of reclassification was 2.62 times greater for students who took the ELA test in English rather than in Spanish. Similarly, the relationship between high performance levels in ELP assessments and probability of reclassification were statistically significant and positive. In regard to student characteristics, Mavrogordato (2012) found that EL students from families with low SES, whose primary language was Spanish, who received special education services, and who had disciplinary infractions, were significantly less likely to be reclassified as English proficient. Furthermore, the probability of reclassification was found to be highest when students were in the third grade. Finally, upon including school context, Mavrogordato (2012) reported that EL students attending schools having higher concentrations of EL students had lower likelihoods of reclassification.

The importance of Mavrogordato's (2012) study in the research on reclassification lies in its being the first conducted using a large dataset in a state other than California. This state, Texas, also has one of the highest proportions of EL students. Thus, the findings are important in determining whether similar patterns emerge despite differing states' education systems.

Another discrete event history analysis on EL students was conducted by Slama (2012) on students who started as kindergarteners in the school year 2002-03 in Massachusetts public schools. The analytic sample included 5353 students, and Slama (2012) followed this cohort for eight years until the end of the seventh grade. Her study yielded four important findings. First, "the majority of the 2002 kindergarten EL cohort was reclassified in Massachusetts schools by third grade" (pg. 40). Second, by the end of the seventh grade, 17% of the sample still had LEP

status. Third, Slama (2012) reported that 22% of EL students were retained in grade at some point over the eight-year period. Finally, Slama (2012, pg. 41) found that “more than half of reclassified students scored below proficient on statewide English language arts and mathematics assessments in elementary- and middle-school grades.”

Slama's (2012) study contributed to the pool of findings from another state than California researching reclassification. As did Mavrogordato (2012), she also found that EL students who started school in the U.S. in the first grade or earlier were more likely to be reclassified in their third year. However, a substantial proportion of EL students required a longer period of time to be reclassified. Also, finding that EL students performed at lower proficiency levels on statewide standardized tests after reclassification was important because it raised the question of the extent to which EL students' academic proficiency during early elementary grades could predict their future academic success. One limitation of this study was the failure of the study to account for student level demographic characteristics in the models. Such an approach would most likely have accounted for a part of the variability in reclassification exhibited by the data.

Drawing on the studies discussed so far, some overall patterns emerge. First, nearly all studies suggest that most EL students are reclassified into mainstream classroom between three and six years and that reclassification rates peak at the third year. This is understandable because, particularly for students who enrolled in schools either in kindergarten or first grade, the third grade is the first time that they are able to participate in ELA assessment. Thus, only by the third year do the school-based teams have one or two years of results from necessary academic assessments to evaluate the capabilities of students to participate in English-only classrooms. Second, almost all studies indicate that higher achievement results are strong predictors of

reclassification, another unsurprising finding because ELP and ELA assessments are the two reclassification requirements with clear proficiency requirements for reclassification eligibility. Third, in studies where reclassification is studied based on student ethnicity or primary language background, EL students with Spanish backgrounds are almost always significantly less likely to be reclassified as English proficient. More research is required to determine why this subgroup of EL students in particular suffers from late reclassification. Among possible causes, some particular cognitive demands associated with learning English could make learning it more difficult for native Spanish speakers or some socio-demographic characteristic such as poverty could interfere to a relatively greater extent with this group's ability to acquire English than with other groups'. Another possible reason might be the high levels of school segregation with other Spanish speakers. This may limit students' opportunity to practice English or diminish the need for excelling in English since majority of the students that they are in communication with are also Spanish speakers. Finally, the last two studies established that students who received special education, who were retained in a grade, and who had experienced disciplinary problems were less likely to be reclassified. These findings are not surprising, as these students can be considered already at high risk with respect to several educational outcomes.

2.4 Summary

This chapter has reviewed significant events and processes that characterize EL students' experiences in public school systems (i.e., federal laws that define EL policies and the processes that schools follow to identify, place, and reclassify EL students). Following this review, the chapter focused specifically on the literature describing challenges in the reclassification of EL students into mainstream classrooms, in particular the effect of policies and processes related to the reclassification. One example of such an effect is the influence of accountability

requirements on school decisions related to timing of EL reclassification. Also cited were general concerns regarding the appropriateness of using statewide standardized assessments to determine how well EL students could be expected to function in mainstream classrooms. Additionally, insufficient knowledge on second language acquisition by local decision makers could lead to differences in reclassification decisions. Thus, although this study defined the reclassification event as a dichotomous variable (i.e., 0 = student had LEP status, 1 = student exited LEP status), this review made it clear that a reclassification decision is not straightforward and involves many gray areas.

The chapter concluded by reviewing previous studies on time required for EL students to exit LEP status and on student and school characteristics related to likelihood of exiting LEP status. From this review, a general consensus emerged: Reclassification rates were highest in the third year following identification as LEP but decreased as students progressed through grades. Also, EL students who had low achievement scores, were retained in a grade, or were from Hispanic backgrounds were found to be significantly less likely to be reclassified into mainstream classrooms. All the findings described above aided in identifying student and school characteristics that should be included in the study's analyses.

Chapter 3. Methodology

This chapter details the methodological aspects of the study, beginning with its design and plausible threats to internal validity. The next section describes data sources, sampling strategies, the outcome of interest, and the variables used during the analysis. The last section describes the analytic approaches that will be employed to examine the effect of *City Connects* on EL students' exiting LEP status, including preliminary descriptive summaries, estimation of propensity score weights, discrete event history models, and sensitivity analysis.

3.1 Research Design

3.1.1 Study design.

In scientific research, random assignment is considered the "gold standard" in causal inference (Shadish, Cook, & Campbell, 2002). Given a sufficiently large sample size, with random assignment, treatment and control groups are expected to be probabilistically equivalent with respect to both unmeasured and measured variables (Rubin, 1974; Shadish et al., 2002). Random assignment, therefore, reduces the possibility that treatment groups differ in a systematic way (Rubin, 1974).

However, in educational research often times random assignment of subjects into treatment and control groups is not possible due to practical, ethical, or political reasons (Rubin, 1974). In such cases, quasi-experiments are more feasible, with the caveat that possible prior differences between treatment groups can pose a threat to internal validity (Bryk & Weisberg, 1977). The research employed to study the *City Connects* intervention has a quasi-experimental design, i.e., schools for whom the study obtained data were not randomly assigned to the *City Connects* intervention (Walsh et al., 2014). Participation in the study was determined as a consequence of district-level interest in the *City Connects* (Walsh et al., 2014). Also, note that

random assignment by itself is not a guarantee neutralizing the bias resulting from other design and implementation aspects of the study (Ginsburg & Smith, 2016). For example, Ginsburg and Smith (2016) identified nonselection bias in randomized control trials due to weaknesses in implementation fidelity and inadequate time given to implementation for the effects to be observed. This certainly could be a threat to internal validity in quasi-experimental studies as well. Thus, in estimating the effect of the intervention, examining plausible threats to internal validity will be critical, and discussion of design elements and statistical controls that can be introduced into analyses to improve the credibility of the estimated causal effects is necessary (Shadish et al., 2002).

3.1.2 Threats to internal validity.

Mill (1843) suggests that arguments such as “A causes B” have to meet certain criteria in order to be considered causal. Shadish et al. (2002, pg. 6) summarize these as “1) A (the cause) must occur before B (the effect), 2) the cause is related to the effect, and 3) there is no alternative explanation for the effect other than the cause.” The third criterion is concerned with threats to internal validity and can be strengthened by eliminating possible causes that could lead to the conclusion of “the relationship between A and B is not causal ... [and] could have occurred even in the absence of the treatment” (Shadish et al., 2002, pg. 54). Based on review of the applicable literature, Shadish et al. (2002) outline eight main threats to internal validity, which are discussed below along with their implications for this study.

3.1.2.1 Ambiguous temporal precedence. This threat relates to whether the independent variable, the one assumed to represent the cause, occurred before the outcome variable, which is a result of the observed effect (Shadish et al., 2002). This threat can be addressed by administering the treatment prior to measuring the outcome. Thus, for this study, one of the

sample restrictions is the requirement of enrollment into the BPS by the first grade at the latest, thereby allowing for the treatment represented by the *City Connects* intervention to be in effect from the beginning of a student's BPS schooling. The same restriction also applies to students in comparison schools to ensure that all students' school enrollment histories begin within the BPS so as to eliminate any confounding effect resulting from attending a school in another school district.

3.1.2.2 Selection. Selection bias occurs if treatment and control samples differ in systematic ways (Rubin, 1974). If differences are related to the observed variables, then the bias is considered to be explicit (overt), whereas if they are related to unobserved variables, it is considered to be hidden (Shadish et al., 2002). If selection bias exists, attributing an observed treatment effect solely to the treatment is problematic, as it can also be due in part to these systematic differences. The strongest way to minimize the effect of this bias is through random assignment of subjects to treatment conditions, a design feature that should be built into the research. However, as mentioned earlier, random assignment is not always possible in educational research, as is indeed the case for *City Connects*.

While hidden bias is difficult to measure and account for, explicit bias can be reduced by statistical strategies such as matching. This study will use propensity score weights to account for selection bias on the observed variables (Rosenbaum & Rubin, 1983b). To estimate propensity score weights, this study used binary logistic regression and modeled the probability of being assigned to treatment or control groups conditional on pre-treatment characteristics. The procedures involved in estimating propensity score weights are discussed further in Section 3.3.2.

3.1.2.3 Maturation. This threat is likely to affect a study's internal validity because some

consequences of the passage of time, in this case biological growth of participants during the course of the study, could be confounded with the treatment effect (Shadish et al., 2002). Since data for this study encompass students' schooling from first through fifth grades, this threat constitutes a plausible threat to validity. To reduce the maturation effect, this study ensured that students included in the analysis from treatment and comparison schools are of the same age, so that time has a similar effect on their growth.

3.1.2.4 Regression to the mean. This threat becomes a concern when individuals are assigned to one of the treatment conditions due to their extreme scores on a variable or construct (Shadish et al., 2002). Measurement of constructs always includes some measurement error, and, thus, individuals do not always score similarly on the same construct over time. For instance, suppose a low score on a reading test qualified a student to be included in a treatment based on reading ability. Should the result from the first test underestimate the student's reading ability, an increase in the post-test cannot be attributed solely on the treatment effect. For the current study, this threat is not an issue since assignment to the treatment represented by *City Connects* participation did not occur as a result of specific characteristics of students or schools.

3.1.2.5 Attrition. This threat is plausible if attrition causes the treatment and control samples to be systematically different from each other (Shadish et al., 2002), thus confounding systematic differences in the samples with the treatment effect. For the current study, the attrition bias is not a plausible threat to internal validity because in survival analysis attrition is considered as a type of censoring. Censoring refers to cases for which the target event was not observed before the end of the data-collection process (Allison, 1982; Guo, 2009; Singer & Willett, 1993). In other words, although attrition may happen for some cases, they are still kept in the sample for survival analysis, as the purpose of survival analysis is to estimate the distribution of time to a focal

event. In this study, the sample definition required that students to be enrolled in one of the *City Connects* or comparison schools by the start of first grade at the latest and it allowed only for right-hand and independent censoring (see details on Section 3.3.3.1).

3.1.2.6 Testing. Administration of the same or parallel tests more than once over time may influence participants' performance due to such reasons as familiarity or practice (Shadish et al., 2002). These factors can constitute threats to internal validity. The *City Connects* intervention does not include pre- and post-testing of study participants, and thus this threat is not relevant to the current study.

3.1.2.7 Instrumentation. In this study, students' LEP status is the outcome and whether they are current or former LEP students is clearly indicated in the data set. Thus, instrumentation is not a plausible threat to the internal validity of this study.

3.1.2.8 History. Shadish et al. (2002, pg. 56) describe the history threat as "events that occur between the beginning of the treatment and the post-test that could have produced the observed outcome in the absence of that treatment." For this study, one such history threat is possible. In 2010, the BPS reached a settlement agreement with the Office for Civil Rights of the U.S. Department of Education concerning the improper and misidentification of students who were ELs (U.S. Department of Justice, 2010). It was determined that since 2003 more than 8000⁵ students were not provided with the EL services to which they were entitled, either due to improper identification or misidentification. As a result of this settlement, in 2010 the BPS agreed to reclassify those students as LEP and to provide them with EL services. Moreover, this

⁵ Out of this investigation, it was determined that the misidentification happened in two ways. First, while approximately 4000 students were initially identified as students who were EL, they were inappropriately opted out of EL services. Second, an additional 4300 students were never identified as students who were EL, as these students were not tested in all of the four language domains: listening, speaking, reading, and writing.

settlement led to the improvement of identification practices and procedures as well as EL instruction across the district.

The scenario described above may affect this study in two ways. First, these improperly or misidentified students may have been unevenly distributed across *City Connects* versus non-*City Connects* schools. However, considering that this event occurred at the district level, there is a low likelihood that those students are disproportionately distributed across *City Connects* and non-*City Connects* schools. Second, improvements in EL instruction, identification, placement, and reclassification may affect the outcome of exiting LEP status in the positive direction. However, since these changes affected the whole district, there is no reason to think that students in the *City Connects* schools were affected more than students in the non-*City Connects* schools.

3.1.2.9 Assumptions for causality. In addition to examining threats to internal validity, it is important that causal treatment effects, which are estimated through statistical analysis, are unbiased. For unbiased estimation of causal treatment effects, Rubin (1986, 1990) and Rosenbaum & Rubin (1983) list two assumptions that research designs should meet: 1) the stable-unit-treatment value assumption (SUTVA) and 2) the strongly ignorable treatment assignment assumption. Below are short descriptions of these assumptions and their implications for this study.

3.1.2.9.1 SUTVA. Rubin (1986, pg. 961) describes two conditions to be met for SUTVA: 1) “the value of Y for unit u when exposed to treatment t will be the same no matter what mechanism is used to assign treatment t to unit u ” and 2) “the value of Y for unit u when exposed to treatment t will be the same no matter what treatments the other units receive.” Moreover, these two assumptions should hold for all $u = 1, \dots, N$ and all $t = 1, \dots, T$. Violations of SUTVA can happen in two ways: neighborhood effects and treatment group non-adherence.

- **Neighborhood effects.** Violations of SUTVA can happen when study participants share the same environment because the treatment received by some students/schools may affect the response given by other students/schools. In this study, treatment and comparison schools are all in one school district, and thus, they share the same BPS district and the neighborhood. Having the intervention take place in some schools may affect other schools through shared connections. For example, teachers in comparison schools may talk to teachers in treatment schools and learn about the services and resources that *City Connects* provides, to which they may participate independent of the *City Connects* (e.g. community partnerships with organizations like Big Brothers Big Sisters would be available to anyone who applies). Likewise, students in treatment schools who have siblings, cousins, or close friends attending comparison schools may have a peer effect on one another, either positive or negative. Therefore, the treatment given to some schools may affect the treatment received by other schools. In addition, regardless of shared connections, it is very common for schools in BPS to have community partnerships with a few organizations or some type of support services available. Thus, adopting some of the same services that are also available through the *City Connects* can be considered as business as usual for schools in BPS. However, *City Connects*’ theory of action is much more than just providing resources and services. The *City*

Connects provides a system that makes it possible to serve all students within a school. Each student is evaluated by trained site coordinators at least once a year and is provided with a targeted set of enrichment and prevention services based on students' strengths and weaknesses. Also, students are monitored throughout the year to assess the progress and if necessary are provided with new ones. Thus, even though non-*City Connects* schools may participate in some of the same services that are also available through *City Connects*, those services may not be as effective since the match between services and students' needs, as well as the monitoring systems, are not in place as they are in the *City Connects*.

- **Treatment group non-adherence.** Students' mobility between schools should be examined with respect to group non-adherence. The *City Connects* treatment does not impede students' transfer from one school to another. Thus, transfers from *City Connects* schools to non-*City Connects* schools can occur and vice-versa. If students transferred from a non-*City Connects* school to a *City Connects* school, then these students will be flagged as pre-treatment students for the period that they were in non-*City Connects* schools and then flagged as *City Connects* students once they were in one of the treatment schools. Since the sample definition will require that *City Connects* students be enrolled in one of the treatment schools by first grade at the latest, the sample definition automatically excludes pre-treatment students from the analysis. In the opposite case, students that started in a *City Connects* school but then transferred to a non-*City Connects* school will always be considered *City Connects* students in the context of the study and will therefore also be automatically excluded from the control sample. Because of these measures, the final analytic sample will be unlikely to include students who might pose a group non-adherence threat. Thus, interference between treatment and control students within this study is implausible.

3.1.2.9.2 Strongly ignorable treatment assignment assumption. Rosenbaum and Rubin (1983) describe this assumption as the independence of treatment assignment and potential outcomes, given the observed and unobserved pre-treatment variables. In randomized studies, every individual has a chance of receiving each treatment and which treatment they are given does not depend on potential outcomes. Simply put, this implies that treatment assignment is strongly ignorable given a vector of pre-treatment variables. However, this assumption is likely violated when randomization is not used as the assignment mechanism. For this study, schools were assigned to treatment conditions as whole units, and so the assignment mechanism was non-random. Consequently, this study used propensity score weights to balance scores so as to estimate the probability of assignment to a treatment group given observed pre-treatment variables (Rosenbaum & Rubin, 1983b) (see section 3.3.2 for further details).

3.2 Data Description

3.2.1 Data sources.

To estimate the *City Connects* treatment effect, this study drew data from two sources: 1) the BPS student database and 2) the Student Support Information System (SSIS) (City Connects, 2014). Since *City Connects*' first implementation, its evaluation team has received full data from BPS on students for each academic year. This data includes students' demographic characteristics: gender, ethnicity, poverty status (as indicated by free- or reduced-priced lunch status), LEP status, Special Education Status, number of school changes, and retention records. Also included in this database are students' academic data, in the form of report card scores, and scores and proficiency levels from statewide standardized assessments. *City Connects* also generates its own database, called SSIS, for purposes of record keeping, measuring fidelity of implementation, and conducting research, and this database includes information about students'

treatment status, number of years in a *City Connects* school, cohort membership, and tier level⁶.

These two databases are linked every year to incorporate new data on each student and to generate the *City Connects* ' longitudinal database.

3.2.2 Sample.

Table 3.1 depicts the longitudinal data structure of the *City Connects* intervention with the study cohorts shown on the top row and the school year on the far left column. Each cohort is labelled by the school year that the cohort's students were in kindergarten. For example, Cohort 2003 entered the study as kindergartners during 2003 and attended school through the fifth grade either at a *City Connects* or a non-*City Connects* school. Also, each school year is named according to the Fall term of that academic year, so, for example, School Year 2003 encompasses Fall 2003, Winter 2004, and Spring 2004.

⁶ The *City Connects* ' site coordinators evaluate each child in four domains: 1) academic, 2) social/emotional/behavioral, 3) health and 4) family. They assign each student to one of four tier levels based on the intensity of needs and strengths across these four domains(City Connects, 2014). The four tier levels are: Tier 1) Strengths and minimal needs, Tier 2a) Strengths and mild needs, Tier 2b) Strengths and moderate needs, and Tier 3) Strengths and severe needs.

Table 3-1: *City Connects Longitudinal Data Structure.*

Cohort	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12
School																	
Year																	
2001	5	4	3	2	1	K2	K1	K0									
2002	6	5	4	3	2	1	K2	K1	K0								
2003	7	6	5	4	3	2	1	K2	K1	K0							
2004	8	7	6	5	4	3	2	1	K2	K1	K0						
2005	9	8	7	6	5	4	3	2	1	K2	K1	K0					
2006	10	9	8	7	6	5	4	3	2	1	K2	K1	K0				
2007	11	10	9	8	7	6	5	4	3	2	1	K2	K1	K0			
2008	12	11	10	9	8	7	6	5	4	3	2	1	K2	K1	K0		
2009		12	11	10	9	8	7	6	5	4	3	2	1	K2	K1	K0	
2010			12	11	10	9	8	7	6	5	4	3	2	1	K2	K1	K0
2011				12	11	10	9	8	7	6	5	4	3	2	1	K2	K1
2012					12	11	10	9	8	7	6	5	4	3	2	1	K2
2013						12	11	10	9	8	7	6	5	4	3	2	1

This study utilized a subset of the *City Connects* dataset, comprising students who were identified as LEP in one of 12 cohorts: Cohorts 2001 through 2012. Membership in the treatment group was defined as any student who has ever attended a *City Connects* school by the start of first grade at the latest. The comparison group was defined as students who were enrolled in the BPS since the start of the first grade but who have never attended a *City Connects* school. Additionally, students with severe special educational needs requiring instruction in substantially separate classrooms were excluded from the sample, although other special education students remained. Finally, this study used propensity score weights at the student level. Any students or schools having missing values in any of the variables used to estimate propensity score weights were excluded from the analyses. This is because logistic regression analysis in Statistical Package for the Social Sciences (SPSS) uses only the cases with complete data, and thus, does not produce the probabilities if cases have missing values for the variables included in the analysis. Before the estimation of propensity scores at the student level, the sample sizes were 3152 and 12871 students for the *City connects* and non-*City Connects*, respectively. The final analytic samples had 2745 *City connects* and 11062 non-*City Connects* students, once the students who were missing report card scores were eliminated from the sample.

3.2.3 Outcome variable.

Table 3.2 displays a portion of the student-level data set as a hypothetical example. The data were longitudinal in nature and included information regarding student ID, grade, dummy discrete-time variables corresponding to grades, and the outcome with respect to exiting LEP status.

The outcome variable, which is displayed in the last column, indicates outcome of the event “exiting LEP” for each student and is coded dichotomously per grade: 0 = student was

LEP, 1 = student exited LEP status. In this data set, students have one row of data that corresponds to each grade. As shown in the table, the number of rows per student will vary depending on the last grade for which the student was censored or the grade at which the student experienced the target event.

Table 3-2: *Sample Student-Level Data Set.*

Student ID	Grade	Discrete-Time Dummy Variables					Event Indicator
		Gr1	Gr2	Gr3	Gr4	Gr5	Exiting out of LEP status
X	1	1	0	0	0	0	0
X	2	0	1	0	0	0	0
X	3	0	0	1	0	0	0
X	4	0	0	0	1	0	0
X	5	0	0	0	0	1	0
Y	1	1	0	0	0	0	0
Y	2	0	1	0	0	0	0
Y	3	0	0	1	0	0	1
Q	1	1	0	0	0	0	0
Q	2	0	1	0	0	0	0
Q	3	0	0	1	0	0	0
Q	4	0	0	0	1	0	1

3.2.4 Student-Level variables.

At the student level, models included four types of variables: student demographic characteristics, measures of degree of disadvantage, measures of academic performance, and *City Connects* related variables. Table 3.3 presents and describes these variables in detail.

Table 3-3: *Description of Student-Level Variables.*

Variables	Descriptions	Reference Group or Value Ranges
Students' Demographic Characteristics		
Gender	A dummy variable indicating gender.	1= Male 0= Female
Ethnicity/Race	Four dummy variables indicating subjects' race: Black, Asian, Hispanic, and Others.	White
Measures of Degree of Disadvantage		
Special Education Status	Two dummy variables indicating subjects' special education classification: SPED 2 (student was pulled out no more than 25% of regular education), and SPED 3 (student was pulled out no more than 60% of regular education)	SPED 1 (never needed or regular education with modifications)
Lunch Price Status	Two dummy variables indicating students' lunch price status: Free lunch (student receives free lunch), and reduced-price lunch (student receives reduced-price lunch)	Full-Price Lunch
Foreign Born	A dichotomously coded variable indicating whether the subject was born outside of the United States.	1= born outside of the U.S. 0= born inside the U.S.
Measures of Academic Performance		
Mobility	The total number of school moves subjects experienced within the BPS system before 1 st grade	0 to 1
Retention	The total number of retentions in grade subjects experienced within the BPS system before 1 st grade.	0 to 1
Academic Performance	Students' Standardized Report Card Scores from fall of Grade 1 in Math, Reading, Writing, Effort, Behavior, and Work-Habits	-3 to +3
City Connects Related Variables		
City Connects_Dummy	A dichotomously coded variable indicating subjects' treatment group membership.	1= City Connects student, 0= Comparison student.

3.2.5 School-Level variables.

This study used school level characteristics reported in National Center for Education Statistics (NCES) Common Core of Data (CCD) for BPS. The schools were restricted to those that identified themselves as regular public schools⁷ and were required to have grades 1 through 5. Because schools started implementing the *City Connects* treatment in different years and to ensure that pre-treatment school characteristics came from the same school years, school level matching were done on two sub-samples of the *City Connects* schools: Sub-sample 1 comprised five *City Connects* schools that adopted the intervention for the first time in school year 2001-2002, and sub-sample 2 comprised the four *City Connects* schools that adopted the intervention for the first time in school year 2007-2008. The school level matchings were carried out separately on these two samples as well as the subsequent discrete-time event history analysis. For each sub-sample, the prior three years of data from NCES CCD were averaged across the years to establish the variables on which the matching was conducted. That is, NCES CCD school years 1998-1999, 1999-2000, and 2000-2001 were used for sub-sample 1, and NCES CCD school years 2004-2005, 2005-2006, 2006-2007 were used for sub-sample 2. Once school level datasets for BPS were established from NCES CDD then they were merged with the *City Connects* master dataset to bring in the *City Connects* dummy variable indicating schools' treatment status (i.e. *City Connects* vs. non-*City Connects*) and average school report card scores from the fall of Grade 1 for each study year. In the *City Connects* master dataset, the earliest available school year data is 2001-2002 to estimate the average school report card scores for the fall of Grade 1. Thus, for sub-sample 1, rather than prior three years, the average school scores

⁷ In NCES CCD, regular public schools are defined as “A public elementary/secondary school that does not focus primarily on vocational, special, or alternative education” (Keaton, 2012, pg. B-4)

from 2001-2002 were used. For sub-sample 2, three prior years were available to estimate the average report card scores for each school. That is: school years 2004-2005, 2005-2006, and 2006-2007 were used to estimate the average school report card scores for the fall of Grade 1.

Table 3.4 presents and describes these variables from NECS CCD in detail.

Table 3-4: *Description of School-Level Variables.*

Variables	Descriptions	Reference Group or Value Ranges
School Size	Total number of students in school	
% Free Lunch	Percent of students eligible to participate in the Free Lunch Program	0-100%
% Reduced Lunch	Percent of students eligible to participate in the Reduced-Price Lunch Program	0-100%
% Asian	Percent of students identified themselves as Asian	0-100%
% Black Students	Percent of students identified themselves as Black	0-100%
% Hispanic Students	Percent of students identified themselves as Hispanic	0-100%
% Other	Percent of students identified themselves as other	
Academic Performance	Average Standardized Report Card Scores in a given school for Fall of Grade 1 during 2001-2012 in Math, Reading, Writing, Effort, Behavior, and Work-Habits	-3 to +3
City Connects_Dummy	A dichotomously coded variable indicating school's treatment status.	1= City Connects school. 0=Comparison school.

3.3 Analytic Strategy

3.3.1 Stage one: Preliminary descriptive analyses.

This section presented descriptive summaries of four types of variables by treatment group: 1) student demographic characteristics, 2) measures of degree of disadvantage, 3) measure of pre-treatment academic performance, and 4) outcome variable. Additionally, independent *t*-test analyses were performed to investigate the extent to which treatment groups differed with respect to pre-treatment variables.

3.3.2 Stage two: Estimation of propensity score weights and school level matching.

To answer research question 2, two different sample balancing approaches were taken. First, propensity score weights were estimated at the student level using the big analytic sample, and then, incorporated into the two level logistic regression model as level-1 weights (i.e. student level). Second, using a subset of the *City Connects* schools from the analytic sample, school level matching were established between *City Connects* and non-*City Connects* schools. The reason was to ensure pre-treatment characteristics came from the same school years since schools started implementing the *City Connects* intervention in different years. Because school level matching resulted in a smaller analytic sample, there were not enough schools to conduct a two-level logistic regression analysis. Thus, one-level logistic regression analysis at the student-level were carried out for the discrete-time event history analysis. Since one-level analysis did not allow us to use school level propensity score weights, student level propensity score weights were used instead. Once the comparable non-*City connects* schools were identified, then within the sub-sample of schools, student level propensity score weights were re-estimated and incorporated into the one-level analysis. The reason for employing school level matching was to allow for the actual assignment level of the *City Connects* treatment to be accounted for. While this approach decreased the total number of schools in the analysis, it allowed us to examine if the greater likelihood of exiting LEP status in the *City Connects* schools were due (at least in part) to pre-existing differences in school characteristics. In the following paragraphs, the methods for the estimation of propensity score weights and school level matching are described in detail.

In random assignment, individuals have known conditional probabilities of being assigned to treatment or control conditions, and, thus, the assignment process does not depend on

any pre-treatment variables (observed or unobserved) (Rosenbaum & Rubin, 1983b; Rubin, 1990). Also implied is that individuals' treatment assignments and potential outcomes are conditionally independent given pre-treatment variables (Rosenbaum & Rubin, 1983b). However, for studies in which individuals are not randomly assigned to treatment conditions, selection bias remains a threat to internal validity (Shadish et al., 2002). For example, in the *City Connects* case, students may have chosen to attend one of the treatment schools for specific reasons, thus leading to systematic differences in the student's pre-treatment measures, both in observed and unobserved variables, as well as differences in outcomes.

Rosenbaum and Rubin (1983) argue that utilizing propensity scores, which are balancing scores that estimate the probability of being assigned to the treatment group given observed pre-treatment variables, reduces the bias in the estimated treatment effect. Since propensity scores are generated using observed pre-treatment variables, this method reduces the explicit bias and some part of the hidden bias if the unobserved variables are correlated with the observed variables.

To estimate propensity scores, this study used binary logistic regression to model the probability of being assigned to a treatment versus a control group conditional on observed pre-treatment variables (Guo & Fraser, 2010). For each student, the propensity score were thus defined as follows:

$$P(Y = 1) = \frac{e^{\beta X}}{1 + e^{\beta X}}$$

where Y indicates the treatment assignment (i.e., *City Connects* vs. non-*City Connects*) and \mathbf{X} is the set of observed pre-treatment variables. For estimating propensity score weights at the student level, the variables listed in Table 3-3 were used. The decision on which variables to use

in estimating propensity score weights were made based on the What Works Clearinghouse (WWC) guidelines (U.S. Department of Education, 2013, 2014b).

In this study, average treatment effect on the treated (ATT) were estimated and used. In randomized control trials, average treatment effects (ATE) weights and ATT weights are equivalent because it is assumed that treated population is not systematically different than the overall population. However, in quasi-experimental studies where assignment into the treatment requires meeting some selection criteria, treated subjects may differ than the overall population. Because the *City Connects* intervention was not assigned at random to the schools, it was plausible that the *City Connects* schools differed in some ways from the comparison schools and may not be representative of the population of elementary schools in BPS. Thus, this study deemed more appropriate to use the ATT weights instead of ATE in the analyses.

To estimate ATT propensity score weights, this study used the method suggested by Guo and Fraser (2010), where 1 was assigned to every treated student and $[P/(1-P)]$ was assigned for comparison students. Before incorporating propensity score weights into the analysis, first the distribution of the weights were examined to determine whether there were outliers. Second, the covariate balance was evaluated. This study used standardized bias (Harder, Stuart, & Anthony, 2010) to examine the differences in pre-treatment variables before and after the propensity score weighting at the student-level.

For matching at the school level, because the number of schools adopting the *City Connects* intervention in the same year was small, this study used two sub-samples of the *City Connects* schools and conducted the same matching and discrete event history analysis on the two sub-samples separately. The purpose of the replication was to examine whether results held in both samples, thus strengthening the evidence of *City Connects*' treatment effects. Sub-sample

1 comprised six *City Connects* schools that adopted the intervention for the first time in school year 2001-2002, and sub-sample 2 comprised the four *City Connects* schools that adopted the intervention for the first time in school year 2007-2008.

For matching, this study used school characteristics reported in NCES CCD. Schools were matched on variables for schools' total enrollment, racial composition, proportion of students qualifying for free- and reduced-price lunch, and mean school achievement in Grade 1 report card scores from the BPS dataset. For each sub-sample, the prior three years of data from NCES CCD were averaged across the years to establish the variables on which the matching was conducted. That is, NCES CCD school years 1998-1999, 1999-2000, and 2000-2001 were used for the sub-sample 1, and NCES CCD school years 2004-2005, 2005-2006, 2006-2007 were used for the sub-sample 2.

This study used optimal matching to identify the set of comparable non-*City Connects* schools. In optimal matching, first the pairwise distances between all schools are estimated based on the pre-treatment variables used in the logistic regression predicting the probability of being a *City Connects* versus a non-*City Connects* school. Next, with an iterative process, the optimal matching algorithm assigns a control case to a treatment case by minimizing the average distance among all the matched cases (Rosenbaum, 1989). In other words, matching is not done sequentially as in greedy matching (or nearest neighborhood matching), where a treatment case is assigned to a match of minimum distance, and then, are removed from the pool of unmatched cases. Thus, in the next step, the best match is determined based on shortest distance between the remaining pool of treatment and comparison cases. In contrast, optimal matching can re-consider a match that has been already made and revise that match in order minimize the average distance among all the matches (Rosenbaum, 1989). To describe this iterative process better, consider the

following matrix displaying the distances between two hypothetical treatment and comparison cases:

		Treatment	
		1	2
Comparison	1	0	5
	2	5	20

In greedy matching, treatment case 1 would be assigned to comparison case 1 since the distance between the two is 0. Thus, treatment case 2 is left with no choice but to be matched with comparison case 2, where the distance between the two is 20. However, in optimal matching, treatment case 1 would be matched with comparison case 2, and treatment case 2 would be matched with comparison case 1, thus, leading to a smaller average distance among all the matches (Rosenbaum, 1989).

Using the optimal matching method, a treatment school can have more than one matched comparison school. In this study, for each City Connects school two matched comparison schools were identified. Thus, for sub-sample 1 there were six treatment and 12 comparison schools, and for sub-sample 2 there were four treatment and eight comparison schools. Because the total number of schools was small for each matched sub-sample (i.e. 18 schools for sub-sample 1 and 12 schools for sub-sample 2), the discrete event history analysis were carried out as one level analysis. Once the matched comparison schools were identified, propensity scores were re-estimated at the student level within the new analytic samples and were incorporated into the one level analyses.

3.3.3 Stage three: Discrete event history analysis to estimate the *City Connects* effect.

To estimate the effect of receiving the *City Connects* intervention on exiting LEP status, this study employed a discrete event history analysis. Methods of survival analysis answer questions that involve the timing of events as well as whether occurrence of the event differs as a result of characteristics of research participants (Allison, 1982; Guo, 2009; Singer & Willett, 1993; Yamaguchi, 1991). In general, the survival analysis data structure includes a dependent variable indicating whether the event of interest has occurred during the course of the research and a variable indicating the time elapsed until the occurrence of that event (Guo, 2009; Singer & Willett, 1993). If the time variable is measured in discrete-time intervals rather than continuously, such survival research is called discrete event history analysis (Allison, 1982; Singer & Willett, 1993). In using a discrete event history analysis, this study treated the event indicator of exiting LEP status as the outcome variable.

3.3.3.1 Censoring. Censoring is a key concept in survival analysis and refers to cases for which the target event was not observed before the end of the data-collection process (Allison, 1982; Guo, 2009; Singer & Willett, 1993). Censoring has three basic forms: right-hand censoring, left-hand censoring, and independent-censoring. This study allowed only for right-hand and independent-censoring.

Right-hand censoring occurs when the target event was not observed by the end of data collection although the participant was followed for the entire duration of the study (Guo, 2009; Singer & Willett, 1993). This study defined the study window as five years, from first through fifth grades. Thus, students who did not experience the event of exiting LEP status at the end of the data collection process were right-hand censored (see Figure 3.1, pg. 41, Line C). Right-hand

censoring indicates that those students' time to the event was greater than five years or that it never happened.

Independent-censoring, on the other hand, refers to cases for which the starting point exists; however, data collection is terminated before the end of data collection for reasons other than the occurrence of the event of interest (Guo, 2009; Singer & Willett, 1993). For example, students who have data starting from first grade but who have moved out of the school district before fifth grade without experiencing the target event represent independent-censored cases (see Figure 3.1, Line E).

Left-hand censoring occurs when we do not know the starting point of the possibility of experiencing the target event (Guo, 2009; Singer & Willett, 1993). The requirement that students be enrolled in the BPS by the start of first grade at the latest (Figure 3.1, Line D) excluded such cases.

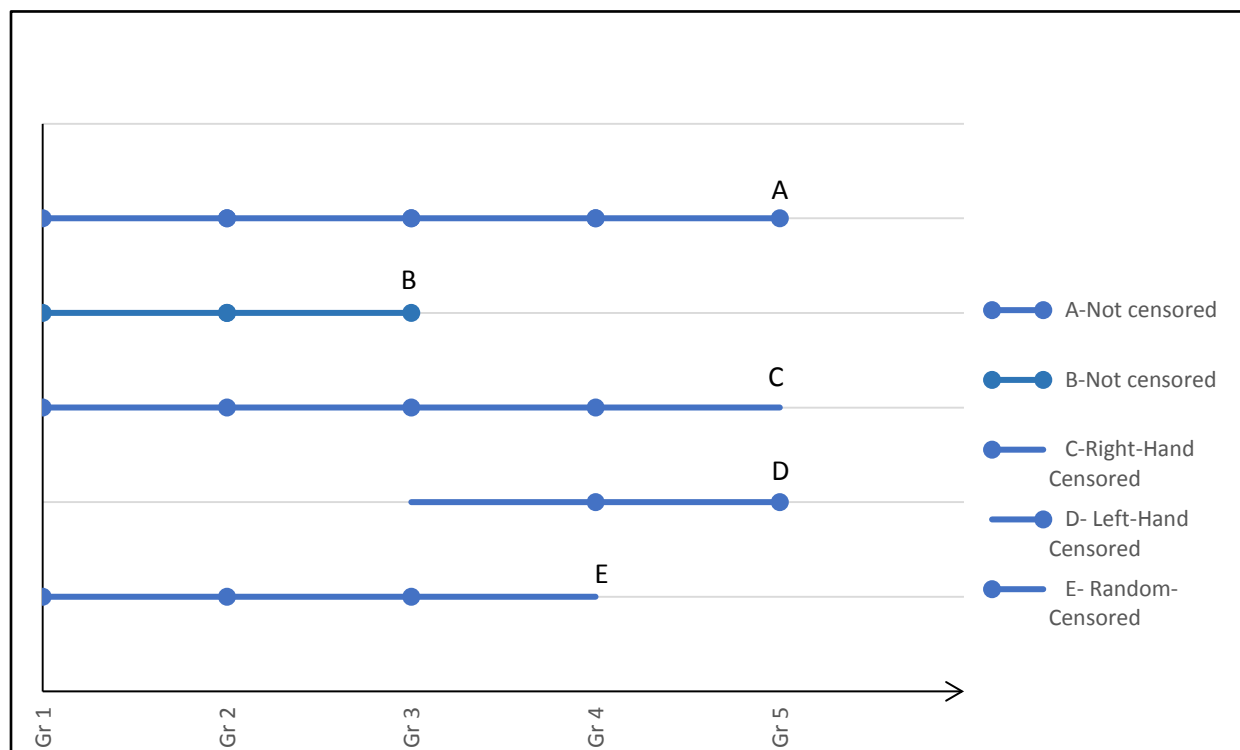


Figure 3-1: Illustration of censoring.

3.3.3.2 Estimating the parameters of the discrete-time hazard model.

Following the work of Singer and Willett (1993), the hazard function can be described as a conditional probability, with time represented as contiguous discrete-time intervals indexed by j ($j = 1, \dots, J$):

$$T = (0_0, t_1], (t_1, t_2], \dots, (t_{j-1}, t_j]$$

Given that the event of interest has not occurred prior to the beginning of time period j , the conditional probability that a randomly selected person will experience the event during time period j can be expressed as follows (Allison, 1982; Singer & Willett, 1993):

$$h_j = \Pr[T = j | T \geq j] \quad (3.1)$$

In estimating the magnitude of the discrete-time model parameters, the study's purpose was to investigate the dependence of the discrete-time hazard on the covariates of interest, such as student background characteristics (e.g., gender or free- or reduced-price lunch status, etc.). Thus, to introduce heterogeneity into the statistical model, we defined P predictors, Z_p , where $p = 1, 2, \dots, P$, and each Z_p corresponds to a specific observed variable (Singer & Willett, 1993). For example, Z_{1ij} could represent student gender and Z_{2ij} each grade's achievement results with i indexing the individual and j indexing the corresponding time period. For a variable that varies over time (e.g., achievement results at each grade), the value of Z_{2ij} would most likely change for each time period that it was observed. On the other hand, for gender, the value of Z_{1ij} would remain constant through each time period for each individual. As shown in Equation 3.2, Equation 3.1 can be re-written to represent the probability that the event will occur in time period j given the covariates and that the individual did not experience the event prior to time period j :

$$h_{ij} = \Pr\{T_i = j | T_i \geq j, Z_{1ij} = z_{1ij}, Z_{2ij} = z_{2ij}, \dots, Z_{pij} = z_{pij}\} \quad (3.2)$$

Considering that h_{ij} are (conditional) probabilities, Cox (1972) describes the statistical model for h_{ij} in logistic form where h_{ij} depends on dummy variables that indicate time periods and on a set of covariates of interest. This population discrete-time hazard model can be written as follows:

$$h_{ij} = \frac{1}{1 + e^{-(\alpha_1 D_{1ij} + \alpha_2 D_{2ij} + \dots + \alpha_J D_{Jij}) + (\beta_1 Z_{1ij} + \beta_2 Z_{2ij} + \dots + \beta_P Z_{Pij})}} \quad (3.3)$$

Where

- j indexes $j = 1, 2, \dots, J$ time periods
- i indexes $i = 1, 2, \dots, I$ individuals
- p indexes $p = 1, 2, \dots, P$ predictors
- $[D_{1ij}, D_{2ij}, \dots, D_{Jij}]$ are a sequence of dummy variables indexing time periods, and J refers to the last time period observed for anyone in the sample,
- $[\alpha_1, \alpha_2, \dots, \alpha_J]$ are the intercept parameters capturing the baseline level of hazard in each time period, and
- $[\beta_1, \beta_2, \dots, \beta_P]$ are the slope parameters describing the effects of the predictors on the hazard function (Singer & Willett, 1993, pg.166).

Equation 3.3 can be transformed into log odds of h_{ij} as follows:

$$\log\left(\frac{h_{ij}}{1 - h_{ij}}\right) = (\alpha_1 D_{1ij} + \alpha_2 D_{2ij} + \dots + \alpha_J D_{Jij}) + (\beta_1 Z_{1ij} + \beta_2 Z_{2ij} + \dots + \beta_P Z_{Pij}) \quad (3.4)$$

Equation 3.4 establishes that the dummy time variables and the set of covariates are linearly related to the log odds of the hazard function (Singer & Willett, 1993). This model contains multiple intercepts $\alpha_1, \alpha_2, \dots, \alpha_J$, one per time period, instead of a single one. These intercepts describe the population baseline logit-hazard function when all covariates equal zero (Singer & Willett, 1993). Next, estimation of the parameters $\alpha_1, \alpha_2, \dots, \alpha_J$ and $\beta_1, \beta_2, \dots, \beta_P$ using the maximum likelihood (MLE) method is presented.

3.3.3.3 MLE of hazard model parameters.

In this section, the ML method is used to estimate the values of $\alpha_1, \alpha_2, \dots, \alpha_j$ and $\beta_1, \beta_2, \dots, \beta_p$. Singer and Willett (1993) described the likelihood function for a discrete event history as the product of censored and uncensored cases. That is:

- 1) uncensored individuals (denoted by $c_i = 0$): the probability that the event occurs in time period j_i (j_i is the time period when the event occurred) but not in periods 1 through $j_i - 1$ can be written as the product of terms expressing conditional probability for each time period:

$$\Pr\{T_i = j_i\} = \Pr\{T_i = j_i | T_i \geq j_i\} * \Pr\{T_i \neq j_i - 1 | T_i \geq j_i - 1\} * \dots * \Pr\{T_i \neq 1 | T_i \geq 1\} \quad (3.5)$$

This can be re-expressed in terms of h_{ij} :

$$\Pr\{T_i = j_i\} = h_{ij_i}(1 - h_{i(j_i-1)}) (1 - h_{i(j_i-2)}) \dots (1 - h_{i2})(1 - h_{i1}) \quad (3.6)$$

A compact version of Equation 3.6 can be written as follows:

$$\Pr\{T_i = j_i\} = h_{ij_i} \prod_{j=1}^{j_i-1} (1 - h_{ij}) \quad (3.7)$$

- 2) censored individuals (denoted by $c_i = 1$): similarly, the probability that the event occurs after time period j_i can be described as the product of conditional probabilities per time period:

$$\Pr\{T_i > j_i\} = \Pr\{T_i \neq j_i | T_i \geq j_i\} * \Pr\{T_i \neq j_i - 1 | T_i \geq j_i - 1\} * \dots * \Pr\{T_i \neq 1 | T_i \geq 1\} \quad (3.8)$$

which can be re-expressed in terms of h_{ij} :

$$\Pr\{T_i > j_i\} = (1 - h_{ij_i})(1 - h_{i(j_i-1)}) (1 - h_{i(j_i-2)}) \dots (1 - h_{i2})(1 - h_{i1}) \quad (3.9)$$

Similar to Equation 3.7, a simpler version of Equation 3.9 can be written as follows:

$$\Pr\{T_i > j_i\} = \prod_{j=1}^{j_i} (1 - h_{ij}) \quad (3.10)$$

Using Equations 3.7 and 3.10, the likelihood function becomes the product of probabilities for censored (Equation 3.11) and uncensored (Equation 3.12) individuals (Singer & Willett, 1993) as follows:

$$L_i = \prod_{i=1}^n [\Pr\{T_i = j_i\}]^{1-c_i} [\Pr\{T_i > j_i\}]^{c_i} \quad (3.11)$$

$$L_i = \prod_{i=1}^n [h_{ij_i}] \prod_{j=1}^{j_i-1} (1-h_{ij})^{1-c_i} \left[\prod_{j=1}^{j_i} (1-h_{ij}) \right]^{c_i} \quad (3.12)$$

Then, the log-likelihood transformation of Equation 3.12 becomes:

$$l = \sum_{i=1}^n [(1-c_i) \log h_{ij_i} + (1-c_i) \sum_{j=1}^{j_i-1} \log(1-h_{ij}) + c_i \sum_{j=1}^{j_i} \log(1-h_{ij})] \quad (3.13)$$

Equation 3.13 can be simplified to obtain Equation 3.14 below (see Appendix A for the simplification steps):

$$l = \sum_{i=1}^n [(1-c_i) \log \left(\frac{h_{ij_i}}{1-h_{ij}} \right) + \sum_{j=1}^{j_i} \log(1-h_{ij})] \quad (3.14)$$

We can also define the outcome of y_{ij} based on whether the individual is censored or not.

If the individual is not censored, then $c_i = 0$, and y_{ij} is equal to 1 only for the last period and 0 for all earlier periods (Singer & Willett, 1993). Similarly, if the individual is censored, then $c_i = 1$ and y_{ij} is equal to 0 for all the time periods, including the very last one (Singer & Willett, 1993). This can be expressed as follows:

$$\sum_{j=1}^{j_i} y_{ij} = (1-c_i) = \begin{cases} 1 & \text{when } c_i = 0 \\ 0 & \text{when } c_i = 1 \end{cases} \quad (3.15)$$

Equation 3.15 can be re-written as the following by multiplying both sides by the same term $\log \left(\frac{h_{ij_i}}{1-h_{ij}} \right)$:

$$\sum_{j=1}^{j_i} y_{ij} \log \left(\frac{h_{ij}}{1-h_{ij}} \right) = (1-c_i) \log \left(\frac{h_{ij_i}}{1-h_{ij}} \right) \quad (3.16)$$

We can then substitute the right side of Equation 3.16 into Equation 3.14, re-arrange terms, and collect like terms to obtain Equation 3.17 below from Equation 3.14 (see Appendix B for the rearrangement of the terms):

$$l = \sum_{i=1}^n \sum_{j=1}^{j_i} [\log h_{ij}^{y_{ij}} + \log(1 - h_{ij})^{(1-y_{ij})}] \quad (3.17)$$

Then, by antilogging Equation 3.17, the likelihood function becomes:

$$L = \prod_{i=1}^n \prod_{j=1}^{j_i} h_{ij}^{y_{ij}} (1 - h_{ij})^{(1-y_{ij})} \quad (3.18)$$

Note that Expression 3.18, the likelihood function for the discrete-time hazard function, is equivalent to the likelihood function representing independent Bernoulli trials with parameters h_{ij} (Allison, 1982; Singer & Willett, 1993). Thus, the probability distribution model that our data follows is equivalent to a Bernoulli distribution. This allows us to estimate the parameters of a discrete hazard function using the methods of standard logistic regression analysis as a function of discrete-time variables and covariates of interest (Allison, 1982; Singer & Willett, 1993). Also, in multilevel modeling, it allows us to treat the longitudinal discrete event history data at the between-student level, which then permits clustering at the school level. Remember that we previously expressed h_{ij} in the form of a logistic function with a set of intercept ($\alpha_1, \alpha_2, \dots, \alpha_j$) and slope ($\beta_1, \beta_2, \dots, \beta_p$) parameters. The likelihood in 3.18, when maximized with respect to these parameters, provides the MLE of the intercept and slope parameters.

3.3.4 Stage four: Modeling strategies for the discrete-time hazard model.

3.3.4.1 Research question 1. The first research question asks “For each grade level, what proportion of students exit LEP status before the next grade in *City Connects* schools and in comparison schools.” To answer this question, I used the most basic method of survival analysis,

the life-table analysis method, to provide initial descriptions of the proportions of students who exited LEP status in *City Connects* and comparison schools. While this method allowed us to compare the LEP exit rates in both *City Connects* and comparison schools, it did not provide an answer as to whether there is a significant difference between exit rates of students in the *City Connects* versus the comparison schools. However, using the Kaplan-Meier method, I examined whether there was a statistical difference in survival functions for exiting LEP status in elementary grades between *City Connects* and comparison groups. This provided the initial evidence on whether the survival distributions differed between the two groups. These analyses were carried out separately for each sample (i.e. the big analytical sample, sub-sample 1, and sub-sample 2). No propensity score weights were incorporated at this time.

3.3.4.2 Research question 2. The second research question asks “To what extent is the *City Connects* intervention associated with students’ likelihood of exiting LEP status while in an elementary grade after adjusting for student characteristics?” This research question was addressed by using two methodological approaches: 1) a discrete-time event history model using one-level logistic regression model as suggested by Singer and Willet (1993) and 2) a two-level logistic regression model that accounts for the nesting structure of the data, where level-1 will contain student level data and level-2 will contain school level data. The following paragraphs describe these two approaches in detail.

Model 1. In this model, a discrete-time event history analysis model was developed using the standard logistic regression method as suggested by Singer and Willet (1993). This model allowed us to examine the association between exposure to the *City Connects* intervention and students’ likelihood of exiting LEP status at any grade through elementary school after

accounting for student characteristics. This analysis was carried on sub-sample-1 and sub-sample-2, respectively. Also, results from the models with ATT weights were reported.

$$\log\left(\frac{h_{ij}}{1-h_{ij}}\right) = \alpha_1 * Grade1_{ij} + \alpha_2 * Grade2_{ij} + \alpha_3 * Grade3_{ij} + \alpha_4 * Grade4_{ij} + \alpha_5 * Grade5_{ij} + (\beta_1 Z_{1ij} + \dots + \beta_P Z_{Pij}) + \beta_{P+1} CCNX_{ij} \quad (3.19)$$

where:

- i indexes $i = 1, 2, \dots, I$ individuals;
- j indexes $j = 1, 2, \dots, J$ time periods;
- $[Grade1_{ij}, Grade2_{ij}, Grade3_{ij}, Grade4_{ij}, Grade5_{ij}]$ are a sequence of dummy variables indexing time periods;
- Notice also that the discrete-time hazard model contains no single stand-alone intercept. Instead the alpha parameters, $[\alpha_1, \dots, \alpha_5]$, act as multiple intercepts, one per time period” (Singer & Willett, 1993 pg. 167);
- $[\beta_1, \dots, \beta_P]$ are the regression coefficients in log odds associated with the P covariates that describe the effects of the predictors on the hazard function;
- $[Z_{1ij}, \dots, Z_{Pij}]$ are P student-level covariates for student i in time period j ;
- β_{P+1} is estimated treatment effect in log odds for the *City Connects* students;
- $CCNX_{ij}$ is the dummy variable indicating the *City Connects* exposure for students, with 1 for *City Connects* and 0 for comparison students;

Model 2. In the second model, a two-level logistic regression model was built to account for the dependency among students from the same schools (Raudenbush & Bryk, 2002). This model was carried out in steps to predict students’ likelihood of exiting LEP status. In the first step, an unconditional model was built, where there was no predictor at level-1 or level-2. This step allowed for the decomposition of the variability in likelihood of exiting LEP status into within- and between-school variance components. The intraclass correlation coefficient (ICC) was estimated using Snijder and Bosker's (1999) latent variable approach for the level-1 model assuming a Bernoulli distribution. In this method, the ICC is computed as

$$\rho = \tau_{00}/(\tau_{00} + \pi^2/3).$$

In the second step, the intercept from the unconditional model was taken out and replaced with the dummy coded discrete-time variables indicating each elementary grade at the student-level. In other words, the intercept was estimated for each discrete-time period one by one instead of as one intercept averaging across all the discrete-time variables. The coefficients of the discrete-time variables captured the mean baseline level of hazard in each time period for all last elementary schools in the big analytic sample. The dummy variable indicating *City Connects* membership was then added at the school-level to predict the coefficients of each discrete-time variable. The magnitude of the *City Connects* dummy variable indicated the extent to which mean hazard of *City Connects* schools differed from the mean hazard of the comparison schools. At this step, the *City Connects* dummy was kept to predict each of the discrete-time variables in order to estimate *City Connects* schools' deviation from the comparison schools regardless of their significance levels.

In the third step, other student level covariates were added to into the model to account for the available variance in the likelihood of exiting LEP status. These student-level variables included: student demographic characteristics, measures of degree of disadvantage, measures of academic performance, and *City Connects* related variables. Because these covariates were strongly related with students' academic success, they were kept in the model regardless of their significance levels. Also, because this study is interested in the school level treatment effects, student level covariates were centered on their grand-mean. With the grand-mean centering, the level-1 intercept becomes the mean across level-2 units adjusted by level-1 covariates (Raudenbush & Bryk, 2002). As the groups of student-level variables added into the model, the dummy variable indicating *City Connects* membership was added at the school-level to predict the coefficients of each of the slope predictors. At this step, those *City Connects* dummy

variables that are statistically significant at the .05 level were retained. Those retained were the ones that significantly predict the within-school slopes. Finally, random effects were examined to determine how much variation is explained in the intercept and slopes with the *City Connects* dummy. If the random effects were not statistically significant, they were fixed, thus not allowed to vary. The results from the model with ATT weights were reported. While the final model took a much simpler form the full statistical model can be expressed as the following:

Level-1 Model:

The Level 1 model equations are as follows:

$$\text{Prob}(\text{EVENT}_{ijk} = 1 \mid \beta_k) = h_{ijk} \quad (3.20a)$$

$$\log[h_{ijk}/(1 - h_{ijk})] = \eta_{ijk} \quad (3.20b)$$

$$\eta_{ijk} = \alpha_{1k} * \text{Grade1}_{ijk} + \alpha_{2k} * \text{Grade2}_{ijk} + \alpha_{3k} * \text{Grade3}_{ijk} + \alpha_{4k} * \text{Grade4}_{ijk} + \alpha_{5k} * \text{Grade5}_{ijk} + \beta_1(Z_{1ijk} - \overline{Z_{1..}}) + \beta_P(Z_{PIJK} - \overline{Z_{P..}}) \quad (3.20c)$$

Level-2 Model:

The Level 2 model equations are as follows:

$$\alpha_{1k} = \gamma_{10} + \gamma_{11}CCNX_k + u_{1k} \quad (3.21a)$$

$$\alpha_{2k} = \gamma_{20} + \gamma_{21}CCNX_k + u_{2k} \quad (3.21b)$$

$$\alpha_{3k} = \gamma_{30} + \gamma_{31}CCNX_k + u_{3k} \quad (3.21c)$$

$$\alpha_{4k} = \gamma_{40} + \gamma_{41}CCNX_k + u_{4k} \quad (3.21d)$$

$$\alpha_{5k} = \gamma_{50} + \gamma_{51}CCNX_k + u_{5k} \quad (3.21e)$$

$$\beta_{1k} = \gamma_{60} + \gamma_{61}CCNX_k + u_{6k} \quad (3.21f)$$

...

$$\beta_{pk} = \gamma_{p0} + \gamma_{p1}CCNX_k + u_{pk} \quad (3.21g)$$

where:

- i denotes $i = 1, 2, \dots, I$ students within elementary schools, J denotes $j = 1, 2, \dots, J$ time periods and J the last time period observed for anyone in the sample, and k denotes $k = 1, 2, \dots, K$ schools for the last elementary school attended;
- $Grade1_{ijk}, \dots, Grade5_{ijk}$ are a sequence of dummy variables indexing time periods;
- $[\alpha_{1k}, \dots, \alpha_{5k}]$ are the regression coefficients in log odds capturing the mean level of hazard in each time period for the last elementary school k ;
- $[\beta_1, \dots, \beta_p]$ are the regression coefficients in log odds associated with the P covariates that describe the effects of the predictors on the hazard function;
- $[Z_{1ijk}, \dots, Z_{pijk}]$ are p student-level covariates for student i in time period j and last elementary school k ;
- $\gamma_{10}, \dots, \gamma_{50}$ are the regression coefficients in log odds indicating the means of the discrete-time variables across all last elementary schools when the CCNX dummy indicator is equal to 0;
- $\gamma_{11}, \dots, \gamma_{51}$ are the regression coefficients in log odds indicating the mean treatment effects in log odds for the last elementary *City Connects* schools;
- $CCNX_k$ is the dummy variable indicating treatment membership for students, with 1 for treatment, and 0 for comparison students;
- u_{0k}, \dots, u_{5k} are the random effects at Level 2 equations

3.3.4.3 Research Question 3. The third research question asks “To what extent do the *City Connects* and non-*City Connects* students differ in their median time to exit LEP status?” To answer this question, I plotted the survival probabilities for both the *City Connects* and comparison schools by setting the covariate values equal to the overall average of the schools in the big sample. In these plots, the point in time that corresponds to .50 survival probability indicated the median time to exit LEP status for each group, i.e., the time by which half of the students exited LEP status. If this median time exceeded the duration of the study, which is five years, I used a linear or parabolic interpolation to estimate the length of time required for students to exit LEP status in *City Connects* versus comparison groups.

3.3.4.4 Research Question 4. The last research question asks “How robust were the estimated treatment effects to the presence of unobserved selection bias?” The purpose of this question was to assess the robustness of the results from this study to possible violation of the assumption of strong ignorability, which Rosenbaum and Rubin (1983) describe as the independence of treatment assignment and outcome given a randomized study’s observed and unobserved pre-treatment variables. In studies where the strong ignorability assumption holds, covariates \mathbf{Z}_{pijk} include all covariates that are related to the response as well as to the treatment-assignment mechanism, denoted as R , such that the two potential outcomes $H_{ijk,R=0}$ and $H_{ijk,R=1}$ for an individual i are conditionally independent given the set of observed covariates (Rosenbaum & Rubin, 1983a):

$$(H_{ijk,0}, H_{ijk,1}) \perp R | \mathbf{Z}_{pijk} \quad (3.22)$$

In such studies, the average treatment effect could be estimated without bias. However, this assumption is likely violated when covariates related to both R and $H_{ijk,R}$ are omitted from \mathbf{Z}_{pijk} .

In her dissertation, Diaconu (2012) adopted a sensitive analysis method proposed by Rosenbaum and Rubin (1983a) with the modifications suggested by Montgomery, Richards, and Braun (1986). To examine sensitivity to unmeasured variables, this study will follow Diaconu’s approach, for which I assumed the existence of a binary unobserved variable that was related to both binary treatment assignment and the binary outcome.

To model sensitivity analysis, “one needs to hypothesize a real but unobserved variable that has a relationship both with the treatment assignment and outcome” (An, 2015, pg. 29), thereby causing selection bias. For this analysis, I assumed that parental involvement is the unmeasured variable U , which *City Connects* does not measure. In other words, I hypothesized

that the treatment assignment becomes strongly ignorable given the set of Z_{pijk} and the unobserved covariate U . The following mathematical expression illustrates this relationship:

$$(H_{ijk0}, H_{ijk1}) \perp R | (U, Z_{pijk}) \quad (3.23)$$

This study made two assumptions in order to introduce selection bias. The first assumption captured the relationship between U and the treatment assignment R : parents who are highly involved with their children's education are more likely to enroll them in a *City Connects* school than in a comparison school. The second assumption quantified the relationship between U and the outcome Y : everything being equal, the likelihood of a student exiting LEP status is larger when parental involvement is high compared to that of a student with relatively less involved parents.

In developing the sensitivity analysis, the first assumption was used to determine the pairs of conditional probabilities that were necessary to introduce a strong bias into the data set. The first assumption can be depicted as a conditional probability, i.e., the conditional probability of parental involvement U taking the value u given assignment to treatment $R = r$, expressed as follows:

$$\pi = \Pr(U = u | R = r) \quad (3.24)$$

The table below illustrates the conditional probability π when both U and R can assume the value of either 0 or 1:

Table 3-5: *The Conditional Probability of Parental Involvement U Given Assignment to a City Connects School.*

Conditional probability (π)		R	
		0	1
U	0	μ_{00}	μ_{01}
	1	μ_{10}	μ_{11}

Given the first assumption (i.e., that parents who are highly involved with their children's education are more likely to enroll their children in a *City Connects* school and less likely to enroll them in a comparison school), the relationship between μ_{11} and μ_{10} can be illustrated as follows:

$$\Pr(U = 1|R = 1) > \Pr(U = 1|R = 0) \quad (3.25)$$

or:

$$\mu_{11} > \mu_{10} \quad (3.26)$$

Similarly, the probability of a student's attending a comparison school given low parental involvement ($R = 0$) is greater than the probability of a student's attending a comparison school given high parental involvement ($R = 1$). These probabilities can be depicted as follows:

$$\Pr(U = 0|R = 0) > \Pr(U = 0|R = 1) \quad (3.27)$$

or:

$$\mu_{00} > \mu_{01} \quad (3.28)$$

The greater the difference between μ_{11} and μ_{10} , that is, $\mu_{11} - \mu_{10}$, (or the greater the difference between μ_{00} and μ_{01} , or $\mu_{00} - \mu_{01}$), the stronger is the bias in the data. To perform the sensitivity analysis, this study employed 10 pairs of μ_{11} and μ_{10} , where $0.2 \leq \mu_{11} \leq 0.8$ and $0.2 \leq \mu_{10} \leq 0.8$, such that $\mu_{11} - \mu_{10}$ increases in increments of 0.15. The variable U was

simulated with a Monte Carlo simulation method by sampling from the conditional distribution of U given R .

Table 3-6: *The 10 Pairs of Conditional Probabilities Used in the Simulation of the Unknown Variable U .*

U	μ_{10}	μ_{11}
u_1	0.20	0.35
u_2	0.20	0.50
u_3	0.20	0.65
u_4	0.20	0.80
u_5	0.35	0.50
u_6	0.35	0.65
u_7	0.35	0.80
u_8	0.50	0.65
u_9	0.50	0.80
u_{10}	0.65	0.80

The second assumption, which was used to introduce bias, captured the relationship between U and the outcome H_{ijk} . With everything else assumed to be equal, the likelihood of exiting LEP status is larger when parental involvement is high ($U = 1$) than when parental involvement is low ($U = 0$). To model this bias, the regression coefficient for U (β_U) was fixed as a positive value when the simulated variable U is introduced as a student level covariate in the HLM model of likelihood of exiting LEP status. Also, note that the magnitude of the regression coefficient β_U was assumed to be the same for both the *City Connects* and the comparison schools. To determine the magnitude of β_U , the magnitude of all available student level covariates similar to parental involvement were examined first and then a value greater than the highest positive regression coefficient was assigned to the parental involvement variable U .

Chapter 4. Results

This chapter presents results from the analyses outlined in Chapter Three. It is organized into five sections. The first section reports results from descriptive analyses for the three samples, the big analytic sample, sub-sample 1, and sub-sample 2. Each of the next four sections answers one of the research questions of this study. In section two, the results from the life-table and Kaplan-Meier analyses are presented for each sample. In section three, first, the results from the baseline equivalence with ATT weights are presented and evaluated, and, second the results from the one-level discrete-time event history models for sub-sample 1 and sub-sample 2 are discussed. These one-level models, which were carried out in steps, accounted for the school level clustering to obtain robust standard errors. This section also includes results from the model evaluation and presents fitted hazard probabilities, survival probabilities, and cumulative hazard rates. After discussing the one-level models, this section reports results for the two-level discrete-time event history models for the big analytic sample using ATT weights. The discussion of the two-level models, which were carried out in steps, includes results from the model evaluation and present the fitted hazard probabilities, survival probabilities, and cumulative hazard rates. In section four, the median time to exit LEP status is determined using the plots of the survival probabilities based on the final model generated by the two-level analysis.

Finally, section five reports the results from the sensitivity analysis using the final model generated by the two-level analysis. The sensitivity analysis introduced an unobserved variable U into the model that was related both to the outcome Y and the treatment assignment mechanism R . Thus, the unobserved variable U was simulated and then used to adjust the two-level model in a procedure that was repeated ten times, each time using a different simulated U .

The resulting treatment effects from the adjusted models are presented, and their evaluation to determine the extent of the impact of a hidden bias on the results of the analysis is discussed.

4.1 Descriptive Analyses

Tables 4-1, 4-2, and 4-3 show the characteristics of LEP students in the *City Connects* and comparison groups at the beginning of Grade 1 for the big analytic sample, sub-sample 1, and sub-sample 2, respectively, before propensity score weighting. For the big sample, there were significantly more Asian students and fewer African American and Hispanic students in the *City Connects* group than in the comparison group. The *City Connects* sample had significantly more LEP students who were identified as SPED 2 and fewer LEP students who were identified as SPED 3. Additionally, the *City Connects* group had significantly more LEP students enrolled in Reduced-Price Lunch Program and fewer LEP students enrolled in the Free Lunch Program. Also, while the *City Connects* had significantly more foreign-born LEP students, LEP students in the *City Connects* group scored significantly lower than their counterparts on all Report Card measures except for Effort from the fall of Grade 1. Finally, there were no statistically significant differences between the two groups in terms of gender, percent retained before grade 1, and percent that changed schools before grade 1.

Table 4-1: *Baseline Student Characteristics by Group Membership for the Big Analytic Sample.*

	City Connects N=2745		Comparison N=11062		
Demographic Characteristics					
%Male	49.1%		48.2%		
Race					
% African American	12.5%		20.9% *		
% Asian	24.9% *		10.7%		
% Hispanic	53.9%		60.5% *		
% Other	1.3%		1.0%		
% White	7.4%		6.9%		
Measures of Degree of Disadvantage					
Special Education					
% non-SPED	87%		85%		
% Regular Education with Modifications (SPED 1)	2.8% *		1.8%		
% Regular Education with no more than 25% out (SPED 2)	7.9%		8.6%		
%Regular Education with no more than 60% out(SPED 3)	2.7%		4.4% *		
Poverty Status					
% Receiving Free Lunch	87.9%		89.8% *		
% Receiving Reduced-Price Lunch	3.9% *		2.9%		
% Receiving Full-price Lunch	8.2%		7.3%		
% Foreign-Born	22.7% *		17.5%		
Measures of Academic Performance					
% Retained Before Gr1	2.2%		2.2%		
% Changed School Before Gr1	0.8%		0.6%		
Standardized Report Card Scores from Fall of Gr1					
	Mean	(SD)	Mean	(SD)	
Z_Math_Gr1	-.06	.99	.00	.99	*
Z_ELA_Gr1	-.07	.99	.01	1.00	*
Z_Writing_Gr1	-.08	.98	.02	1.00	*
Z_Effort_Gr1	.02	.94	.00	1.00	
Z_Work Habits_Gr1	-.08	.96	.01	1.01	*
Z_Behavior_Gr1	-.09	.93	.02	1.01	*

*Statistically more or higher than the other group at $p < 0.05$

In sub-sample 1, which consisted of schools that started the *City Connects* treatment for the first time in school year 2001-2002 and the matched-comparison schools, there were significantly more Asian and White students and fewer Hispanic students in the *City Connects* group than in the matched-comparison group. In this sample, the *City Connects* group had significantly more LEP students enrolled in the Reduced-Price Lunch Program and fewer LEP students enrolled in the Free Lunch Program. Additionally, the *City Connects* group had significantly fewer LEP students identified as SPED 3 and more foreign-born LEP students. Finally, there were no statistically significant differences between the two groups in terms of gender, percent retained before grade 1, percent that changed schools before grade 1, and Report Card measures.

Table 4-2: *Baseline Student Characteristics by Group Membership for Sub- Sample 1.*

	City Connects N=628		Comparison N=2627	
Demographic Characteristics				
%Male	48.7%		48.8%	
Race				
% African American	11.0%		9.3%	
% Asian	21.7% *		15.8%	
% Hispanic	48.9%		68.4% *	
% Other	0.6%		0.6%	
% White	17.8% *		5.8%	
Measures of Degree of Disadvantage				
Special Education				
% non-SPED	87%		86%	
% Regular Education with Modifications (SPED 1)	1.8%		2.1%	
% Regular Education with no more than 25% out (SPED 2)	8.6%		7.8%	
%Regular Education with no more than 60% out(SPED 3)	2.2%		4.0% *	
Poverty Status				
% Receiving Free Lunch	87.4%		89.3% *	
% Receiving Reduced-Price Lunch	4.3% *		2.6%	
% Receiving Full-price Lunch	8.3%		8.1%	
% Foreign-Born	23.4% *		14.7%	
Measures of Academic Performance				
% Retained Before Gr1	3.3%		3.2%	
% Changed School Before Gr1	0.5%		0.3%	
Standardized Report Card Scores from Fall of Gr1				
	Mean	(SD)	Mean	(SD)
Z_Math_Gr1	.01	.99	.00	1.00
Z_ELA_Gr1	.02	1.01	-.01	1.00
Z_Writing_Gr1	.02	1.00	.00	1.00
Z_Effort_Gr1	-.03	.98	.01	1.01
Z_Work Habits_Gr1	-.04	.99	.01	1.00
Z_Behavior_Gr1	.00	.94	.00	1.01

*Statistically more or higher than the other group at $p < 0.05$

In sub-sample 2, which consisted of schools that started the *City Connects* treatment for the first time in school year 2007-2008 and the matched-comparison schools, there were significantly more Asian students and fewer African American, Hispanic, and White students in the *City Connects* group than in the matched-comparison group. The *City Connects* group had significantly fewer LEP students who were identified as SPED 3 and fewer foreign-born LEP students. Additionally, LEP students in the *City Connects* group scored significantly lower than their counterparts on all Report Card measures except for ELA and Effort from the fall of Grade 1. Finally, there were no statistically significant differences between the two groups in terms of gender, percent of students who enrolled in the Free- or Reduced-Price Lunch Program, percent retained before grade 1, and percent that changed schools before grade 1.

Table 4-3: *Baseline Student Characteristics by Group Membership for Sub- Sample 2.*

	City Connects N=536		Comparison N=1005		
Demographic Characteristics					
%Male	50.0%		49.6%		
Race					
% African American	8.6%		24.6%		*
% Asian	70.5% *		26.3%		
% Hispanic	17.5%		35.9%		*
% Other	1.1%		0.9%		
% White	2.2%		12.3%		*
Measures of Degree of Disadvantage					
Special Education					
% non-SPED	91%		86%		
% Regular Education with Modifications (SPED 1)	0.7%		1.6%		
% Regular Education with no more than 25% out (SPED 2)	6.2%		8.4%		
%Regular Education with no more than 60% out(SPED 3)	2.1%		4.2%		*
Poverty Status					
% Receiving Free Lunch	84.7%		81.6%		
% Receiving Reduced-Price Lunch	6.0%		4.9%		
% Receiving Full-price Lunch	9.3%		13.5%		
% Foreign Born	12.9%		17.5%		*
Measures of Academic Performance					
% Retained Before Gr1	0.9%		1.7%		
% Changed School Before Gr1	0.4%		0.7%		
Standardized Report Card Scores from Fall of Gr1					
	Mean	(SD)	Mean	(SD)	
Z_Math_Gr1	-.09	.98	.05	1.01	*
Z_ELA_Gr1	-.06	.98	.03	1.01	
Z_Writing_Gr1	-.12	.92	.06	1.03	*
Z_Effort_Gr1	-.01	.95	.00	1.02	
Z_Work Habits_Gr1	-.10	1.04	.05	.97	*
Z_Behavior_Gr1	-.15	1.12	.08	.92	*

*Statistically More or Higher than the other group at $p < 0.05$

4.2 Research Question One

The first research question concerns the proportions of students exiting LEP status before the next grade in the *City Connects* and comparison schools. To answer this question, two methods were used: 1) the life-table analysis method to provide initial estimates of the proportions of students who exited LEP status (in each grade) in the *City Connects* and comparison schools, and 2) the Kaplan-Meier method to examine whether there was a statistically significant difference in the distributions of time to exit LEP status in elementary grades between the *City Connects* and comparison schools.

Tables 4-4, 4-5, and 4-6 display the life-tables for the big analytic sample, sub-sample 1, and sub-sample 2, respectively. These tables indicate whether and, if so when, LEP students exited LEP status during elementary grades. The columns under “Number who” display the number of LEP students at the beginning of each grade, the number who were censored at the end of the grade, and the number who exited LEP status by the end of the grade. The sixth column shows the proportion of students who were still in LEP status at the end of each year, and the seventh column shows the proportion of LEP students known to be in LEP status at the beginning of the grade who exited LEP status by the end of the grade.

The very last columns of the three tables show the proportions of students exiting LEP status. For each year, these proportions were computed by dividing the number of students who exited LEP status during the grade (i.e., column five) with the risk set, which is the number of students who were LEP at the beginning of the grade (i.e., column three), under the assumption

of independent⁸ censoring. In other words, censored cases (i.e., column 4) were not subtracted from the risk set for that year.

In the case of independent-censoring, time to event for censored and uncensored individuals are considered the same, and thus, counting in the censored individuals in the risk set yields an unbiased estimate of the proportion of students exiting LEP status. For this study, it was not possible to empirically establish that censoring was independent due to incomplete data on when censored students experienced the event. However, because it was imperative for the life-table computations, this study still assumed independent-censoring to be the case. Thus, readers should be cautioned that the results from the life-table analysis might be biased to some extent.

The sixth columns in Table 4-4, 4-5, and 4-6 show the percent of students that were still in LEP status by the end of each grade. Table 4-4 indicates that nearly 49% of *City Connects* students and 59% of comparison schools students were still in LEP status by the end of fifth grade for the big analytic sample. These rates were 50% and 58% for sub-sample 1 and 38% and 57% for sub-sample 2, respectively. The very last columns of the three tables show that the proportions of students exiting LEP status was consistently higher in the *City Connects* group than in the comparison group for each grade.

⁸ Independent-censoring refers to cases for which the data collection is terminated before the end of data collection for reasons other than the occurrence of the event of interest (Guo, 2009; Singer & Willett, 1993). In this study, independent-censoring occurred in two ways: 1) either because students transferred out of the district before the end of fifth grade, or 2) data was cut-off for some students because BPS data available for this study only encompassed school years 2002 through 2013.

Table 4-4: *Life Table for the Big Analytic Sample.*

		Number who			Proportion of	
		Were LEP students at the beginning of the grade	Were censored at the end of the grade	Exited LEP status during the grade	All students who are still in the LEP Status at the end of the grade	Students at the beginning of the grade who exited LEP status by the end of the grade
City Connects	Gr 1	2745	551	13	1.00	0.00
	Gr 2	2181	418	137	0.93	0.06
	Gr 3	1626	316	185	0.83	0.11
	Gr 4	1125	274	254	0.64	0.23
	Gr 5	597	459	138	0.49	0.23
Comparison	Gr 1	11062	1664	44	1.00	0.00
	Gr 2	9354	1321	396	0.95	0.04
	Gr 3	7637	946	653	0.87	0.09
	Gr 4	6038	908	940	0.74	0.16
	Gr 5	4190	3346	844	0.59	0.20

Table 4-5: *Life Table for Sub-Sample 1.*

		Number who			Proportion of	
		Were LEP students at the beginning of the grade	Were censored at the end of the grade	Exited LEP status during the grade	All students who are still in the LEP Status at the end of the grade	Students at the beginning of the grade who exited LEP status by the end of the grade
City Connects	Gr 1	628	114	2	1.00	0.00
	Gr 2	512	85	28	0.94	0.05
	Gr 3	399	53	43	0.84	0.11
	Gr 4	303	43	61	0.67	0.20
	Gr 5	199	149	50	0.50	0.25
Comparison	Gr 1	2627	368	4	1.00	0.00
	Gr 2	2255	334	94	0.96	0.04
	Gr 3	1827	278	134	0.89	0.07
	Gr 4	1415	217	245	0.73	0.17
	Gr 5	953	759	194	0.58	0.20

Table 4-6: *Life Table for Sub-Sample 2.*

		Number who			Proportion of	
		Were LEP students at the beginning of the grade	Were censored at the end of the grade	Exited LEP status during the grade	All students who are still in the LEP Status at the end of the grade	Students at the beginning of the grade who exited LEP status by the end of the grade
City Connects	Gr 1	536	66	3	0.99	0.01
	Gr 2	467	73	37	0.92	0.08
	Gr 3	357	68	54	0.78	0.15
	Gr 4	235	57	85	0.50	0.36
	Gr 5	93	71	22	0.38	0.24
Comparison	Gr 1	1005	155	3	1.00	0.00
	Gr 2	847	119	31	0.96	0.04
	Gr 3	697	88	65	0.87	0.09
	Gr 4	544	69	115	0.69	0.21
	Gr 5	360	300	60	0.57	0.17

While life table analysis allows us to compare LEP exit rates descriptively for the *City Connects* and comparison schools, it does not provide an answer as to whether time to exiting LEP status differed meaningfully for students in the *City Connects* and comparison schools. In order to provide the initial evidence on whether the time to this event differed between the two groups, the Kaplan-Meier method and corresponding log rank test were used. The Kaplan-Meier method, a non-parametric estimation method (i.e., one in which no assumptions are made about the probability distributions of the variables used), can be used in the presence of right-hand and independent-censoring (Kaplan & Meier, 1958). It estimates the survival function by taking the product limit of the conditional probabilities of the event occurring during each discrete-time interval (Kaplan & Meier, 1958). The log rank statistic is used for testing the equality of survival distributions estimated by the Kaplan-Meier method.

Table 4-7 displays the percent of students exiting LEP status and percent of censored cases for each sample by group membership, respectively. The descriptive summaries indicate that the event of exiting LEP status occurred more often in the *City Connects* group than in the comparison group for both sub-samples 1 and 2. In the big analytic sample, the percent of students exiting LEP status were the same for the two groups.

Table 4-7: *Descriptive Summary of Exiting LEP Status and Censored Cases.*

	Total N	Exited LEP		Censored	
		N	Percent	N	Percent
Big Analytic Sample					
Comparison	11062	2877	26%	8185	74.0%
<i>City Connects</i>	2745	727	26%	2018	73.5%
Overall	13807	3604	26%	10203	73.9%
Sum-Sample 1					
Comparison	2627	671	26%	1956	74.5%
<i>City Connects</i>	628	184	29%	444	70.7%
Overall	3255	855	26%	2400	73.7%
Sub-Sample 2					
Comparison	1005	274	27%	731	72.7%
<i>City Connects</i>	536	201	38%	335	62.5%
Overall	1541	475	31%	1066	69.2%

Table 4-8 presents the median time to exit LEP status for the three samples by group membership, respectively. The Kaplan-Meier method, as a non-parametric method, only uses the range of times found in the data to establish the survival distribution. Thus, for the samples where the survival probability does not reach the 0.5 probability, it does not estimate the median time to the event. For this reason, the median time to exiting LEP status was reported as greater than five years for either or both *City Connects* and comparison groups within each sample, when that was the case.

While the median time to exit LEP status in the *City Connects* group was five years for the big sample, greater than five years for sub-sample 1, and four years for sub-sample 2, it was always greater than five years for the comparison group in each sample. Considering the big-sample, which actually encompasses the two sub-samples, the results suggest that it takes five years to exit LEP status for an average student in the *City Connects* group. Or, in other words, about half of the LEP students in *City Connects* elementary schools could start enrolling in mainstream classes no later than the beginning of the sixth grade. Although the results suggest that it takes longer than five years for students in the comparison group to exit LEP status, it is difficult to conclude whether the difference between the two groups is meaningful since the results did not provide enough information about the magnitude of this difference.

Table 4-8 also includes the results of the log rank test. The log rank test tests the null hypothesis that there is no difference in the overall survival distributions between the two groups. The results indicate that the survival distributions were statistically significantly different for the two groups in each sample. In other words, observing that the two groups differed in their survival distributions provided the initial evidence prompting further examination of the time to event of *exiting LEP status* with more advanced modeling with research question 2.

Table 4-8: *Median times for Time to Event and Overall Comparison.*

	Median for Survival Time	Overall Comparison		
	Estimate	Log Rank (Mantel-Cox) Chi-Square	df	Sig.
Big Sample				
Comparison	>5	60.103	1	<0.001
City Connects	5			
Overall	>5			
Sum-Sample 1				
Comparison	>5	9.865	1	0.002
City Connects	>5			
Overall	>5			
Sub-Sample 2				
Comparison	>5	39.929	1	<0.001
City Connects	4			
Overall	>5			

4.3 Research Question Two

The second research question asks “To what extent is the *City Connects* intervention associated with students’ likelihood of exiting LEP status while in elementary school after adjusting for student characteristics?” The aim of this question is to estimate the effectiveness of the *City Connects* intervention on students’ likelihood of exiting LEP status, after controlling for student characteristics. To answer this research question, a series of one-level and two-level discrete-time event history models were built with each sample. However, as noted in Chapter 3, before estimating the treatment effect of the *City Connects* intervention, baseline equivalence for each sample must be established first so that it can be incorporated into the models. In the following sections, first, the results from the evaluation of baseline equivalence are displayed for each sample, respectively. Next, the results from the one-level and two-level discrete-time event history models are presented.

4.3.1 Baseline equivalence.

Table 4-9 presents the standardized bias statistics before and after ATT weighting. The unadjusted standardized bias statistics were estimated by dividing each group mean difference by the treatment group's standard deviation. The weighted standardized bias statistics were estimated by first subtracting the unadjusted mean of the *City Connects* group from the weighted mean of the comparison group and then dividing the resulting value by the standard deviation of the treatment group (Harder et al., 2010). For evaluation of the standardized statistics, this study used the guidelines provided by the WWC. Based on these guidelines: 1) if the differences in any baseline characteristics are greater than 0.25 standard deviations, the groups are considered not equivalent, 2) if the differences are between 0.05 and 0.25 standard deviations, the analysis requires statistical adjustment, and 3) if differences are less than or equal to 0.05, no statistical adjustments are required (U.S. Department of Education, 2013).

In Table 4-9, the red-shaded area indicates the characteristics with a level of covariate imbalance that exceeded the WWC guidelines, the green-shaded area indicates those characteristics requiring statistical adjustment, and areas displaying no shade indicate the characteristics that required no statistical adjustment. An examination of the standardized bias before the ATT weighting makes it clear that many variables were unbalanced. However, for all standardized bias statistics after the ATT weighting, values were less than or equal to 0.05, indicating that the pre-existing differences in observed covariates were balanced between the two groups after the ATT weighting.

Table 4-9: Standardized Bias Statistics Before and After ATT Weighting.

	Big Sample		Sub Sample 1		Sub Sample 2	
	Unadj. Std. Bias	Weighted Std. Bias	Unadj. Std. Bias	Weighted Std. Bias	Unadj. Std. Bias	Weighted Std. Bias
Male	-0.02	0.00	0.00	0.00	-0.01	-0.04
African American	0.25	0.00	-0.05	0.00	0.57	0.02
Asian	-0.33	-0.01	-0.14	0.00	-0.97	-0.04
Hispanic	0.13	0.01	0.39	0.01	0.48	0.03
White	-0.02	0.00	-0.31	-0.02	0.68	0.01
Other	-0.03	0.00	0.00	0.00	-0.02	0.01
SPED 2	0.03	0.00	-0.03	0.01	0.09	-0.04
SPED 3	0.11	0.00	0.12	0.00	0.15	-0.01
Free Lunch	-0.05	0.00	-0.08	-0.02	-0.05	-0.02
Reduced-Price Lunch	0.06	0.00	0.06	0.01	-0.09	-0.05
Foreign Born	-0.12	0.01	-0.20	-0.03	0.14	0.03
Retained Before Gr1	-0.03	0.00	-0.02	-0.01	0.05	-0.02
Changed School Before Gr1	0.00	0.00	-0.01	0.01	0.08	0.00
Z_Math_Gr1	0.07	0.02	-0.01	0.01	0.15	0.04
Z_ELA_Gr1	0.09	0.01	-0.03	0.01	0.10	0.07
Z_Writing_Gr1	0.10	0.01	-0.02	0.00	0.19	0.06
Z_Effort_Gr1	-0.02	0.02	0.03	0.01	0.01	0.09
Z_Work Habits_Gr1	0.09	0.02	0.04	0.00	0.14	0.10
Z_Behavior_Gr1	0.11	0.02	0.00	-0.01	0.21	0.15

4.3.2 Results from model 1 (one-level discrete-time event history models).

To explore the association between the likelihood of exiting LEP status and the *City Connects* intervention, one-level discrete-time event history models were carried out in STATA, using the *vce* (cluster *clustvar*) option for sub-samples 1 and 2, respectively. Because the numbers of schools at level-2 were small, a two-level model was not possible for these sub-samples. For sub-sample 1, there were 18 schools at level-2 (six *City Connects* and 12 matched-comparison schools), and for sub-sample 2, there were 12 schools at level-2 (four *City Connects* and eight matched-comparison schools). With STATA's *vce* (cluster *clustvar*) option, this study

was able to account for the clustering effect and estimate the robust standard errors in the one-level analysis (Rogers, 1993).

The one-level discrete-time event history model was carried out in stages by adding five sets of student-level variables to the model one set at a time. These five sets of variables were discrete-time dummy variables indexing time, students' demographic characteristics, measures of degree of disadvantage, measures of academic performance, and the dummy variable indicating *City Connects* exposure. In this model, no school characteristics were included in the analyses. A likelihood-ratio test was not appropriate when the one-level logistic regression model was adjusted for weights and clustering because, while the observations within each cluster were not independent, the weights do not reflect random sample weights (Korn & Graubard., 1990). Instead, at each stage, the Wald test statistic was used to examine whether the inclusion of the new set of variables sufficiently improved the model fit to compensate for the increase in model complexity (i.e., use of additional degrees of freedom). The Wald test was used to test the following null hypothesis: the regression coefficients of the newly introduced variables are equal to zero. A chi-squared value is generated by the Wald test along with the p -value and the corresponding degrees of freedom. If the p -value is sufficiently extreme (i.e., p -value lower than 0.05), it indicates that the null hypothesis should be rejected because the new set of variables yield a statistically significant improvement in the fit of the model.

Table 4-10 presents the model-building process for sub-sample 1 with the ATT weights. As can be seen in Table 4-10, the results of the Wald hypothesis tests suggest that, in each model, the coefficients for the added variables were statistically significantly different from zero when the other variables in the model were controlled for. Additionally, the magnitude of the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) suggest that

Model 1E has the best fit. As a result, from this series of tests, it was determined that Model 1E is the best-fitting model.

In Table 4-10, the coefficients of each discrete-time dummy variable indicate the shape of the baseline logit hazard function and whether the probability of exiting LEP status increases, decreases, or remains the same over time. To understand what these coefficients mean, we will consider the final model: Model 1E. In this model, the reference students are the ones in the matched-comparison schools whose values of all the variables in the model were set to zero. The coefficient estimates of Model 1E indicate that the logit hazard function for comparison students steadily increases from Grade 1 to Grade 5. We can also transform the coefficients to provide the actual hazard values at each time period. For example, at Grade 3, the transformation of the logit hazard to the hazard value for comparison students is as follows:

$$h_{ij} = \frac{1}{1 + e^{-(\alpha_3 Grade3_{ij})}}$$

$$h_{ij} = \frac{1}{1 + e^{-(-2.45)}} = 0.079$$

Thus, based on the fitted model, of the comparison-school students who had not yet exited LEP status prior to Grade 3, the probability of exiting LEP status was nearly 8% during third grade, where the values of all the student characteristics in the model was set to zero.

To understand the likelihood of exiting LEP status among *City Connects* students, examining the effect of the dummy variable indicating *City Connects* membership was necessary. In Model 1E, the estimate for CCNX coefficient is 0.33 in log odds. This estimate indicates the size of the likelihood differential that existed between the students in *City Connects* and the matched-comparison schools, controlling for all other variables in the model. The positive sign of the estimate indicates that, in every grade, students in the *City Connects* schools were estimated to have a greater probability of exiting LEP status than were the students in

matched-comparison schools, controlling for all other variables in the model. The estimated odds ratio, $e^{(\beta_{CCNX})}$, is 1.393, also indicates that at every grade the estimated odds of exiting LEP status was nearly 1.4 times higher for *City Connects* students than for those in the matched-comparison schools, controlling for all other variables.

Finally, Table 4-10 Model 1E reports regression coefficients investigating the relationship between student characteristics and likelihood of exiting LEP status. To facilitate interpretation, estimated coefficients in log-odds can be transformed into odds ratios. Also, note that, an odds ratio of 1 indicates that the two groups have the same probability of exiting LEP status; odds ratios greater than 1 indicate that a particular group is more likely to exit LEP status; and odds ratios less than 1 indicate that a particular group is less likely to exit LEP status. Along with these, note that when odds ratios are greater than 1, the difference between the odds ratio and 1 represents the difference in the likelihood of reclassification between the two groups. However, odds ratios less than 1 are harder to visualize. For example, odds ratio of 0.2 would mean “0.2 people will experience the event for every one that does not. This translates into one event for every five non-events” (Davies, Crombie, & Tavakoli, 1998, pg.990). Thus, while interpreting odds ratios smaller than 1, this study will divide 1 with the odds ratio in order to reverse the interpretation for the group expected to have higher odds.

Examining fitted Model 1E, we see that girls were 26% (as indicated by 1 divided by boys’ odds ratio of 0.79, $p < .01$) more likely than boys to exit LEP status. Similarly, students who did not receive special education services or received regular education only with modifications were twice more likely to exit LEP than students in SPED 2 (as indicated by 1 divided by SPED 2’s odds ratio of 0.48 ($p < .01$)) and 10 times more likely to exit LEP than students in SPED 3 (as indicated by 1 divided by SPED 3’s odds ratio of 0.10 ($p < .01$)).

Moreover, while Asian students were 68% ($p < .05$) more likely to exit LEP status than their White peers, foreign born students were 38% ($p < 0.01$) more likely to exit LEP status than their U.S. born peers. Finally, students with higher levels of report card scores in math were 38% ($p < .01$), reading 17% ($p < .05$), and effort were 18% ($p < .05$) more likely to exit LEP status compared to their peers who scored at the lower levels.

Table 4-10: *Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Sub-Sample 1, ATT Weighted.*

<i>Variables in the Equation</i>	Model 1A		Model 1B		Model 1C		Model 1D		Model 1E		OR
	B	S.E.	B	S.E.	B	S.E.	B	S.E.	B	S.E.	
Grade 1, α_1	-6.10**	0.46	-6.10**	0.45	-5.54**	0.56	-6.17**	0.54	-6.36**	0.52	0.00
Grade 2, α_2	-2.99**	0.19	-2.98**	0.24	-2.39**	0.33	-3.00**	0.34	-3.19**	0.36	0.04
Grade 3, α_3	-2.34**	0.21	-2.31**	0.28	-1.69**	0.43	-2.27**	0.39	-2.45**	0.38	0.09
Grade 4, α_4	-1.46**	0.09	-1.41**	0.23	-0.75*	0.36	-1.25**	0.34	-1.43**	0.35	0.24
Grade 5, α_5	-1.29**	0.12	-1.21**	0.19	-0.53	0.38	-0.95*	0.37	-1.13**	0.36	0.32
Male, β_1			-0.39**	0.08	-0.35**	0.07	-0.24**	0.09	-0.23**	0.08	0.79
Black, β_2			-0.34	0.33	-0.26	0.30	-0.33	0.28	-0.37	0.28	0.69
Asian, β_3			0.60*	0.25	0.60**	0.21	0.55*	0.22	0.52*	0.22	1.68
Hispanic, β_4			0.09	0.20	0.19**	0.18	0.24	0.17	0.22	0.16	1.25
Other, β_5			-0.28	0.50	-0.43	0.42	-0.71	0.52	-0.81	0.50	0.45
SPED2, β_6					-0.99**	0.12	-0.72**	0.13	-0.73**	0.13	0.48
SPED3, β_7					-3.00**	0.52	-2.36**	0.58	-2.35**	0.60	0.10
Reduced-Price Lunch, β_8					-0.52	0.29	-0.40	0.32	-0.37	0.30	0.69
Free-Price Lunch, β_9					-0.68**	0.32	-0.36	0.29	-0.33	0.27	0.72
Foreign-Born, β_{10}					0.13	0.10	0.32*	0.09	0.32**	0.11	1.38
Mobile before Gr1, β_{11}					-0.38	1.02	-0.28	0.73	-0.28	0.69	0.75
Retained before Gr1, β_{12}					0.08	0.40	0.07	0.40	0.12	0.38	1.13
Z_Math, β_{13}							0.17**	0.05	0.16**	0.06	1.17
Z_Reading, β_{14}							0.30*	0.14	0.32*	0.14	1.38
Z_Writing, β_{15}							0.05	0.09	0.03	0.08	1.03
Z_Effort, β_{16}							0.16	0.09	0.17*	0.08	1.18
Z_WorkH, β_{17}							-0.01	0.09	-0.02	0.09	0.98
Z_Beh, β_{18}							0.14	0.08	0.15	0.08	1.17
CCNX, β_{19}									0.33*	0.17	1.40

*p < .05; ** p< .01

Table 4-10 (Continued): *Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Sub-Sample 1, ATT Weighted.*

Model Fit Statistics	X^2	p	X^2	p	X^2	P	X^2	p	X^2	p
Wald Test	826.19	0.001	46.38	0.001	165.82	0.001	155.41	0.001	3.93	0.001
df	5		5		7		6		1	
AIC	2020.92		1998.80		1969.56		1886.86		1881.57	
BIC	2057.50		2071.96		2093.94		2055.14		2057.16	

*p < .05; ** p < .01

Table 4-11 displays the model building process for sub-sample 2 with the ATT weights. The same analytical steps were applied to sub-sample 2 as to sub-sample 1. In Table 4-11, the results of the Wald tests of the usual null hypotheses suggest that, in each model, the coefficients for the added variables were statistically significant when controlling for the other variables in the model. Additionally, the magnitude of the AIC and BIC values suggest that Model 1E has the lowest value, indicating a better fit. As a result, from this series of tests, it is determined that Model 1E is the best-fitting model.

In Table 4-11, the coefficients of each discrete-time dummy variable indicate the shape of the baseline logit hazard function and whether the probability of exiting LEP status increases, decreases, or remains constant over time. To understand the implications of these coefficients, we will consider Model 1E. In this model, the reference students are the ones in the matched-comparison schools whose values of all model variables were set to zero. The coefficient estimates of Model 1E indicate that the logit hazard function for comparison students steadily increases from Grade 1 to Grade 4 but falls in Grade 5. We can transform the coefficients to provide the actual hazard values at each time period. For example, at Grade 3, the transformation of the logit hazard to the hazard value for comparison students is as follows:

$$h_{ij} = \frac{1}{1 + e^{-(\alpha_3 \text{Grade}_{3ij})}}$$

$$h_{ij} = \frac{1}{1 + e^{-(-3)}} = 0.047$$

Thus, based on the model, of the comparison students who had not yet exited LEP status prior to grade 3, the probability of exiting LEP status was nearly 5% during third grade, where the values of all the student characteristics in the model set to zero.

To understand the exiting LEP status among *City Connects* students, the dummy variable indicating *City Connects* membership was examined. In Model 1E, the estimate for the CCNX

coefficient is 0.38 in log odds. This estimate indicates the size of the difference in likelihoods that existed between the *City Connects* and matched-comparison schools, controlling for all other variables in the model. The positive sign indicates that, in every grade, *City Connects* students were at a greater probability of exiting LEP status than were the students in matched-comparison schools, controlling for all other variables in Model 1E. The estimated odds ratio, $e^{(\beta_{CCNX})}$, is 1.469, indicating that, at every grade, the estimated odds of exiting LEP status was nearly 1.5 times higher for *City Connects* students than for those in the matched-comparison schools, controlling for all else in Model 1E. To estimate the hazard probabilities for *City Connects* students, the same procedure as described above was followed but with the coefficient for *City Connects* substituted into the expression estimating h_{ij} . For example, in Grade 3, the probability of exiting LEP status was 6.8%, resulting from the following substitution:

$$h_{ij} = \frac{1}{1 + e^{-(\alpha_3 Grade3_{ij} + \beta_{p+1} CCNX_{ij})}}$$

$$h_{Grade1} = \frac{1}{1 + e^{-(-3*1 + 0.38*1)}} = 0.068$$

In other words, while the probability of exiting LEP status was 4.7% for students in the comparison group it was 6.8% for *City Connects* students, after controlling for all the student characteristics in the model.

Finally, this model suggested that girls were 46% (as indicated by 1 divided by boys' odds ratio of 0.69, $p < .05$) more likely than boys to exit LEP status. Similarly, students who did not receive special education services or received regular education only with modifications were 71% more likely to exit LEP than students in SPED 2 (as indicated by 1 divided by SPED 2's odds ratio of 0.59 ($p < .01$)) and 2.9 times more likely to exit LEP than students in SPED 3 (as indicated by 1 divided by SPED 3's odds ratio of 0.34 ($p < .01$)). Moreover, Asian students were

2.5 times ($p < .05$) more likely to exit LEP status than their White peers. Finally, students with higher levels of report card scores in math were 53% ($p < .01$), reading 53% ($p < .05$), and work habits were 45% ($p < .05$) more likely to exit LEP status compared to their peers who scored at the lower levels in these report card scores. These results were similar with sub-sample 1 Model 1E in regard to the variables observed as significant and the direction of the coefficients.

Table 4-11: *Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Sub-Sample 2, ATT Weighted.*

Variables in the Equation	Model 1A		Model 1B		Model 1C		Model 1D		Model 1E		OR
	B	S.E.	B	S.E.	B	S.E.	B	S.E.	B	S.E.	
Grade 1, α_1	-5.49**	0.43	-5.66**	0.68	-5.26**	0.70	-6.62**	0.64	-6.83**	0.59	0.00
Grade 2, α_2	-2.70**	0.22	-2.86**	0.48	-2.45**	0.54	-3.76**	0.45	-3.97**	0.42	0.02
Grade 3, α_3	-1.87**	0.16	-2.01**	0.42	-1.58**	0.49	-2.81**	0.49	-3.00**	0.49	0.05
Grade 4, α_4	-0.58*	0.27	-0.69	0.44	-0.20	0.52	-1.30*	0.56	-1.47**	0.54	0.23
Grade 5, α_5	-1.41**	0.13	-1.46**	0.38	-0.98*	0.46	-1.94**	0.48	-2.10**	0.48	0.12
Male, β_1			-0.45**	0.09	-0.40**	0.10	-0.39*	0.17	-0.38*	0.17	0.69
Black, β_2			-0.30	0.40	-0.09	0.38	0.09	0.41	0.09	0.41	1.10
Asian, β_3			0.49	0.38	0.55	0.31	0.97**	0.31	0.94**	0.29	2.57
Hispanic, β_4			0.10	0.32	0.32	0.38	0.69	0.45	0.68	0.44	1.97
Other, β_5			0.42	0.46	0.49	0.43	0.92	0.51	0.93	0.53	2.53
SPED2, β_6					-0.86**	0.23	-0.53**	0.19	-0.54**	0.17	0.59
SPED3, β_7					-1.67**	0.27	-1.04**	0.33	-1.07**	0.32	0.34
Reduced-Price Lunch, β_8					-0.15	0.31	0.30	0.41	0.30	0.41	1.34
Free-Price Lunch, β_9					-0.59**	0.27	0.17	0.31	0.16	0.30	1.18
Foreign-Born, β_{10}					0.08	0.14	0.31	0.18	0.32	0.17	1.38
Mobile before Gr1, β_{11}					-0.09	0.50	-0.06	0.56	-0.08	0.53	0.93
Retained before Gr1, β_{12}					0.30	0.43	0.37	0.40	0.37	0.41	1.44
Z_Math, β_{13}							0.42**	0.09	0.43**	0.10	1.53
Z_Reading, β_{14}							0.38	0.19	0.43*	0.18	1.53
Z_Writing, β_{15}							-0.12	0.12	-0.14	0.13	0.87
Z_Effort, β_{16}							-0.14	0.10	-0.17	0.09	0.85
Z_WorkH, β_{17}							0.38**	0.14	0.37**	0.13	1.45
Z_Beh, β_{18}							-0.07	0.13	-0.05	0.12	0.95
CCNX, β_{19}									0.38**	0.15	1.47

* p < .05 ; ** p < .01

Table 4-11 (Continued): *Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Sub-Sample 2, ATT Weighted.*

Model Fit Statistics	X^2	p	X^2	p	X^2	P	X^2	p	X^2	p
Wald Test	393.59	0.001	35.47	0.001	158.1	0.001	215.07	0.001	6.36	0.012
df	5		5		7		6		1	
AIC	1871.45		1850.74		1835.57		1714.76		1707.83	
BIC	1904.18		1916.19		1946.83		1865.29		1864.91	

* p < .05 ; ** p < .01

Figures 4-1 and 4-2 present the fitted hazard and survival probabilities for sub-sample 1, respectively. Within each graph, the blue lines show the estimates for *City Connects* students and the red lines the estimates for students in matched-comparison schools. The plot for hazard probability per each grade for the matched-comparison group is constructed by substituting the particular grade's coefficient into the expression for h_{ij} with all other time indicators and student-level characteristics set to zero. For example, in Grade 1, the probability of exiting LEP status, where all the other variables in the model were set to zero, is very small, nearly 0.17% .

$$h_{ij} = \frac{1}{1 + e^{-(\alpha_1 \text{Grade1}_{ij})}}$$

$$h_{\text{Grade1}} = \frac{1}{1 + e^{-(-6.36)}} = 0.0017$$

Similar computations lead to estimates of h_{Grade2} , h_{Grade3} , h_{Grade4} , and h_{Grade5} for comparison students.

To estimate the hazard probabilities for *City Connects* students, the same procedure as described above was followed but with the coefficient for *City Connects* substituted into the expression estimating h_{ij} . For example, in Grade 1, the probability of exiting LEP status was 0.24%, resulting from the following substitution:

$$h_{ij} = \frac{1}{1 + e^{-(\alpha_1 \text{Grade1}_{ij} + \beta_{p+1} \text{CCNX}_{ij})}}$$

$$h_{\text{Grade1}} = \frac{1}{1 + e^{-(-6.36*1 + 0.33*1)}} = 0.0024$$

An examination of Figure 4-1 reveals that the probability of exiting LEP status increased as students progressed from one grade to the next. Notice that while the pattern of hazard is similar for the *City Connects* and comparison group, there is a vertical separation between the two groups. Specifically, this difference is the largest in fifth grade. Also, we see that the

conditional probability of exiting LEP for *City Connects* students was nearly 1.4 times that of students in the matched-comparison schools in each grade.

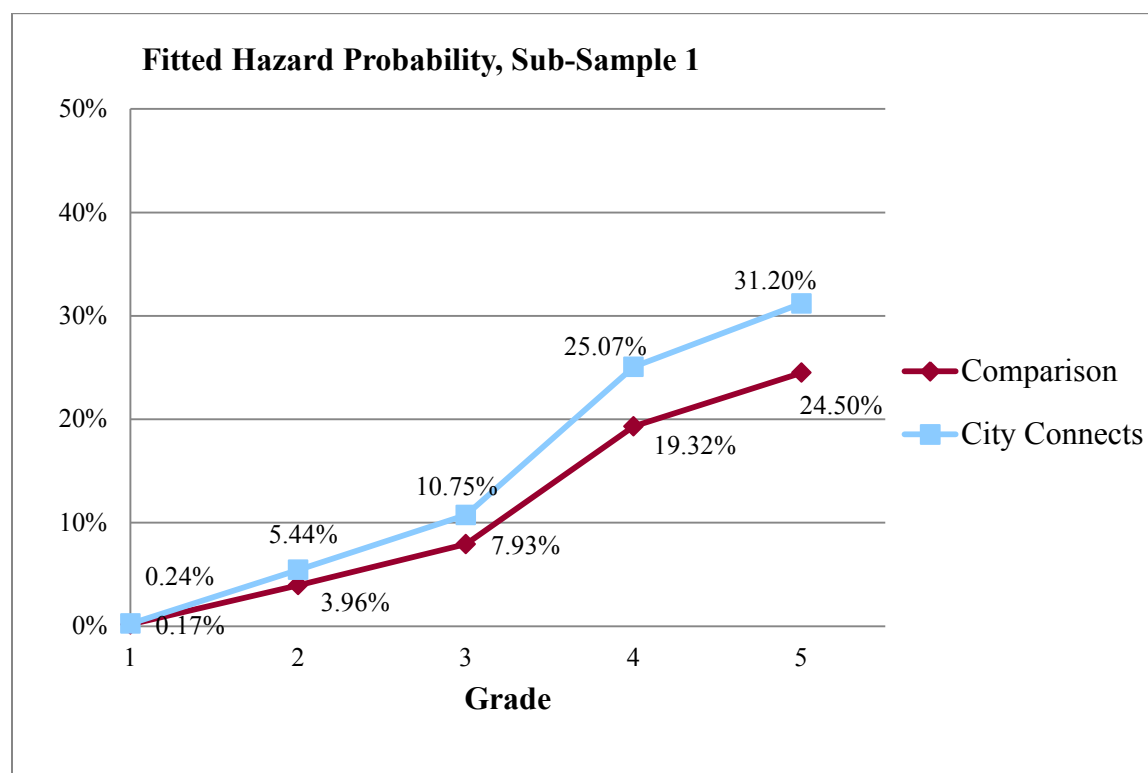


Figure 4-1: Fitted hazard probability for exiting LEP status conditional on student characteristics, sub-sample 1.

Figure 4-2 depicts the estimated survival probability for sub-sample 1. The survival probabilities are estimated by substituting the hazard probabilities into equation 3.10.

$$S_j = \prod_{j=1}^{j_i} (1 - h_{ij})$$

For example, for *City Connects* students, the Grade 1 survival probability was equal to the following:

$$S_{Grade1} = (1 - h_{ij})$$

$$S_{Grade1} = (1 - 0.24)$$

$$S_{Grade1} = 99.76\%$$

For Grade 2, it was equal to the following:

$$S_{Grade2} = (1 - h_{i1}) * (1 - h_{i2})$$

$$S_{Grade2} = 99.76\% * 94.56\%$$

$$S_{Grade2} = 94.33\%$$

Figure 4-2 shows that the percent of students that were still in the LEP status decreased as students progressed from one grade to the next for both groups. In this figure, the point in time corresponding to the .5 survival probability indicates the median time to exit LEP status for each group, i.e., the time by which half of the students had exited LEP status. The estimated median time is 4.7 years for the *City Connects* students and 5.1 years for the matched-comparison students. Thus, results suggest that a typical *City Connects* student exit LEP status during the second half of the fifth grade, while a typical student in the matched-comparison school do so during the start of sixth grade. Hence, while a typical *City Connects* student may start attending mainstream classes at or before the start of sixth grade, a typical student from the matched comparison school may do so at or before the second half of sixth grade. In other words, this difference may translate into a gain of one half of a year in grade in mainstream classes. However, note that this study duration only encompasses the elementary grades. Thus, the plausible changes in school environment and policies from elementary to middle grades may impact the time to exiting LEP status in ways that this study was not able to measure. As a result, the median time to the event that goes beyond five years may be biased to some extent. Lastly, figure 4-2 suggests that by the end of fifth grade, there were nearly 10.4% more students who were still in the LEP status in the matched-comparison schools than in the *City Connects* schools.

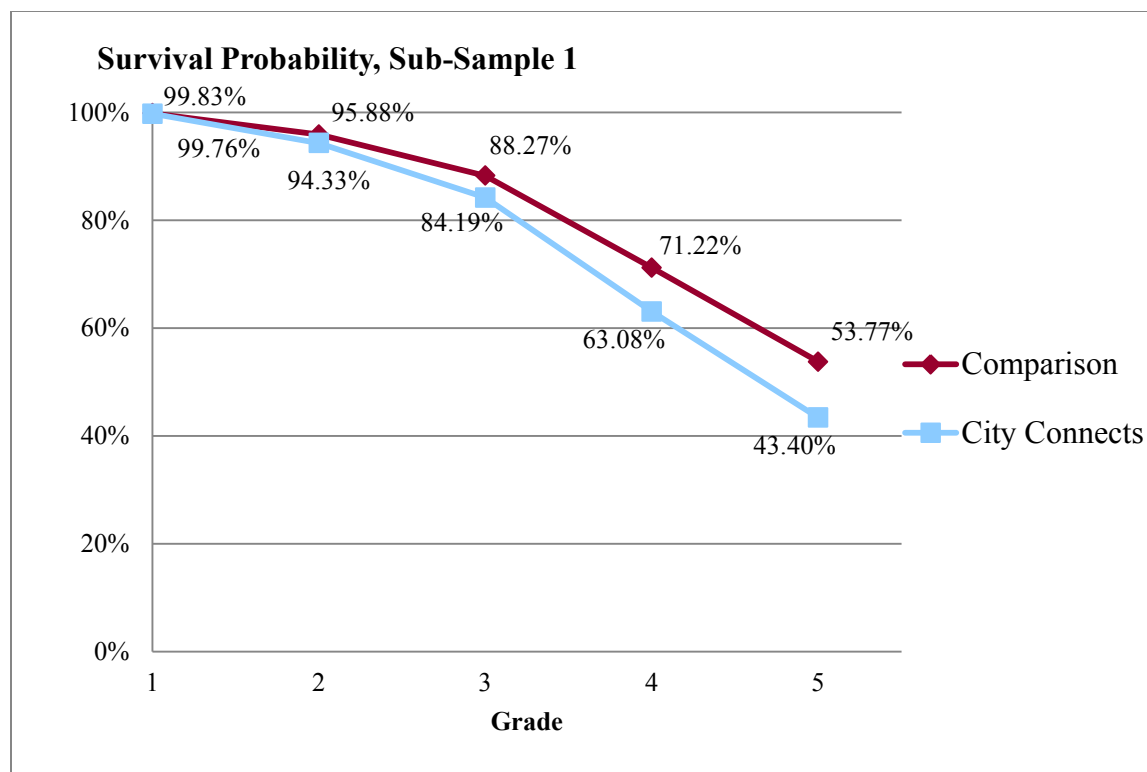


Figure 4-2: Survival probability conditional on student characteristics, sub-sample 1.

The cumulative hazard rate for each group was estimated simply by subtracting the percent of students that were still in the LEP status in Grade 5 in Figure 4-2 from 100%. This yielded values of 56.60% for the students in *City Connects* and 46.23% for the students in matched-comparison schools. In other words, the percent of students who exited LEP status by the end of fifth grade in the *City Connects* schools was nearly 10.4% more than that of students in matched-comparison schools.

Figures 4-3 and 4-4 present the fitted hazard and survival probabilities for sub-sample 2, respectively. Within each graph, the blue lines show the estimates for students in the *City Connects* and red lines for students in matched-comparison schools. The hazard and survival probabilities were estimated using the same procedures as described for sub-sample 1.

Figure 4-3 reveals that the pattern of hazard was similar for the *City Connects* and comparison group. The probability of exiting LEP status first increased at a steep rate for the two groups during Grades 1 to 4 but then declined in Grade 5 for both groups. However, note that there was a vertical separation between the two groups. Specifically, this difference was the largest in Grade 4. Lastly, Figure 4-3 suggests that the probability of exiting LEP status for *City Connects* students was nearly 1.5 times greater than that of students in the matched-comparison schools in each grade.

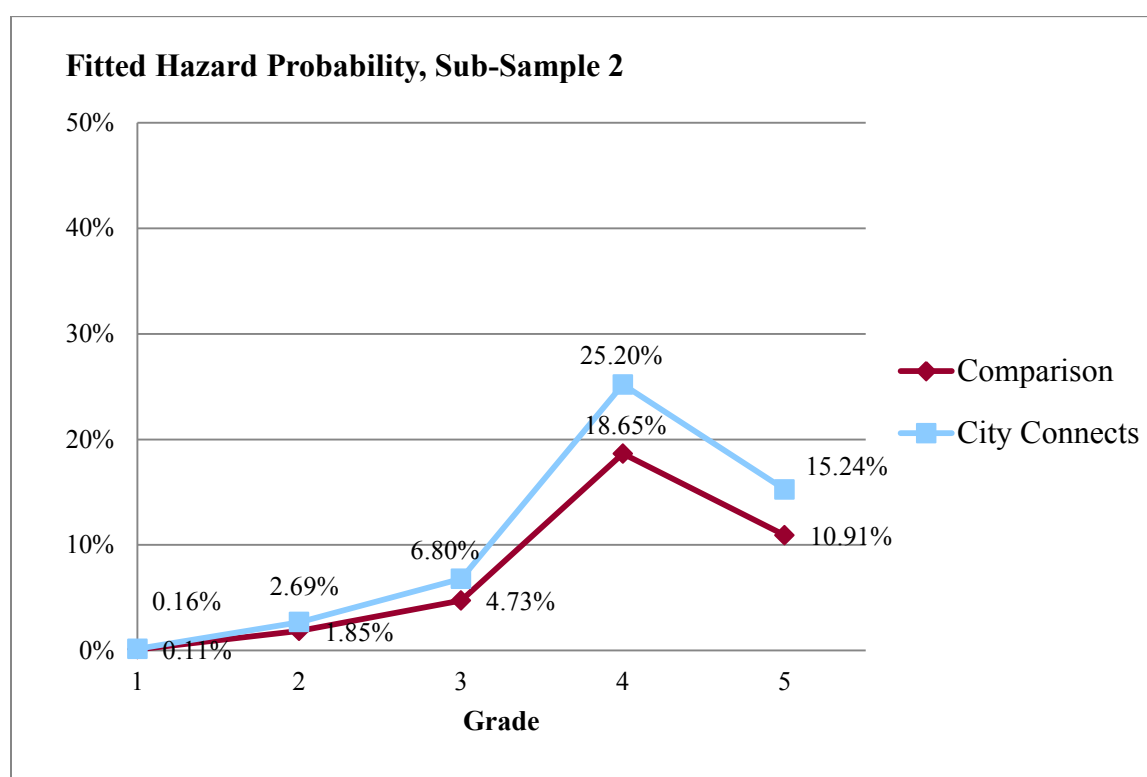


Figure 4-3: Fitted hazard probability for exiting LEP status conditional on student characteristics, sub-sample 2.

Figure 4-4 provides a graphical summary of the two groups' survival probabilities as students in sub-sample 2 progressed from one grade to the next. Note that a survival always trends downward since a participant who experienced the event can never return to the risk set. As indicated by the steeper downward slope, the percent of students exiting LEP status was the

highest in fourth grade. In this figure, the estimated median time is 5.2 years for the *City Connects* students and 5.9 years for the matched-comparison students. Thus, results suggest that a typical *City Connects* student exit LEP status during the first half of the sixth grade, while a typical student in the matched-comparison school do so at the end of sixth grade. This difference may translate into a gain of at least one half of a year in grade in mainstream classes. However, as mentioned earlier, note that the median time to event that goes beyond five years may be biased to some extent. Finally, figure 4-4 reveals that, by the end of the fifth grade, there were nearly 10.3% more students still in the LEP status in the matched-comparison schools than in the *City Connects* schools.

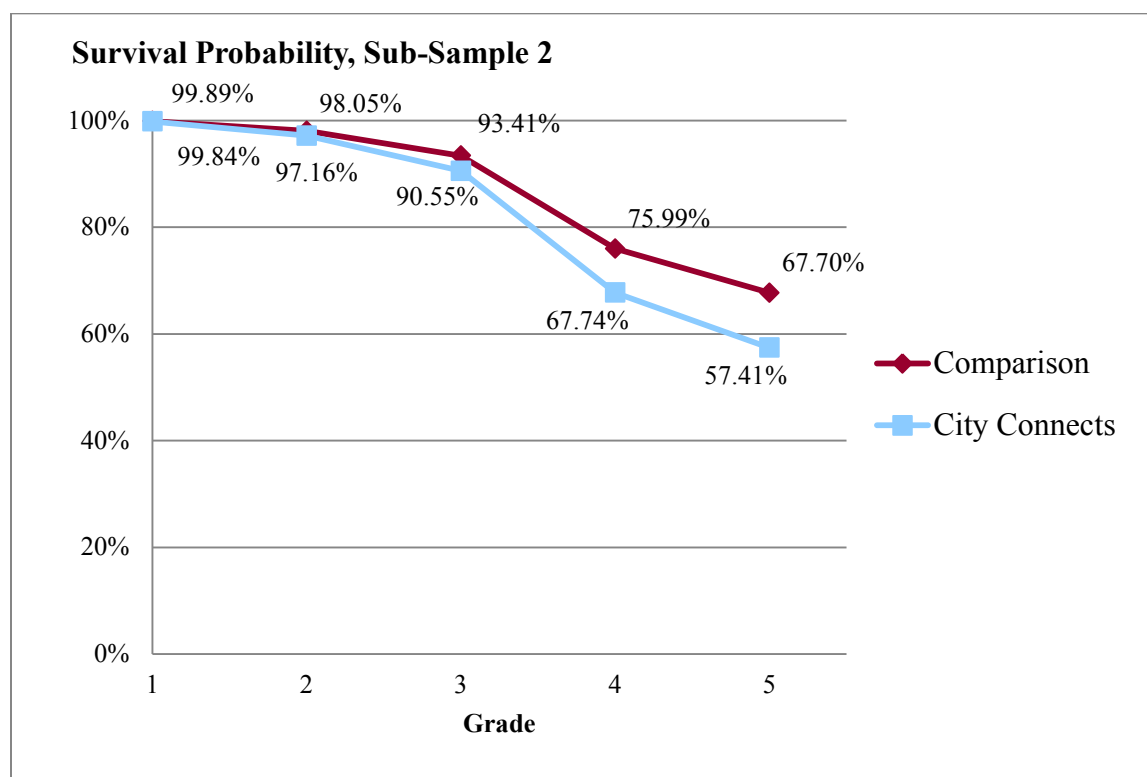


Figure 4-4: Survival probability conditional on student characteristics, sub-sample 2.

The cumulative hazard rate for each group was estimated simply by subtracting the percent of students that were still in the LEP status in Grade 5 in Figure 4-4 from 100%. This

yielded values of 42.59% for the students in *City Connects* and 32.30% for the students in matched-comparison schools. In other words, the percent of students who exited LEP status by the end of fifth grade in the *City Connects* schools was nearly 10.3% more than that of students in matched-comparison schools.

4.3.3 Results from model 2 (two-level discrete-time event history models).

This study used a two-level discrete-time event history model to explore the association between the likelihood of exiting LEP status and the *City Connects* intervention for the big analytic sample using Scientific Software International's Hierarchical Linear and Nonlinear Modeling (HLM) 7 software. The two-level models were carried out in steps. Tables 4-12 through 4-14 present the model-building process for the big sample with the ATT weights.

In the first step, as depicted in Table 4-12, Model 2A, an unconditional model was built, where there were no predictors at level-1 or level-2. The ICC was estimated using Snijder and Bosker's (1999) latent variable approach for the level-1 model assuming a Bernoulli distribution. In this method, the ICC is computed as $\rho = \tau_{00}/(\tau_{00} + \pi^2/3)$. Using this method, the ICC for the unconditional model was thus estimated as 0.04 (i.e., $0.13/(0.13 + \pi^2/3)$), indicating that only 4% of the total variance in likelihood of exiting LEP status was between schools. Although this level of ICC is considered small, this study continued with the multilevel modeling process, because it allowed the effect of the *City Connects* intervention to be tested at the school-level, and so, to examine whether *City Connects* and comparison schools' mean odds of exiting LEP status differed for each discrete-time variable. In addition, it has the capability of providing robust standard errors and account for the data's nesting nature.

In the second step, as depicted in Table 4-12, Model 2B, the intercept from the unconditional model was taken out and replaced with the dummy coded discrete-time variables

indicating each elementary grade at the student-level. In this way, the intercept was estimated for each discrete-time period one by one instead of as one intercept averaging across all the discrete-time variables. At this step, the coefficients of discrete-time dummy variables indicate the mean level of hazard in log odds for all last elementary schools attended in the big analytic sample.

In the third step, as depicted in Table 4-13, Model 2C, the dummy variable indicating *City Connects* membership was added at the school-level. In this step, while the coefficients of $\gamma_{10}, \dots, \gamma_{50}$ correspond to the mean hazard level in log odds for comparison schools, the coefficients of the *City Connects* dummy variable, $\gamma_{11}, \dots, \gamma_{51}$, indicated the extent to which the mean hazard for City Connects schools' deviated from the mean of comparison schools. The *City Connects* dummy variable was retained as a predictor of the discrete-time variables regardless of its statistical significance.

In the fourth step, as depicted in Table 4-13 Model 2C through Table 4-14 2F, sets of student level covariates were added to into the model one by one to account for the available variance in the likelihood of exiting LEP status. These student-level variables included: student demographic characteristics, measures of degree of disadvantage, measures of academic performance, and *City Connects* related variables. Also, because this study is interested in the school level treatment effects, these variables were centered on their grand-mean. With the grand-mean centering, the level-1 intercept becomes the mean across level-2 units adjusted by level-1 student characteristics (Raudenbush & Bryk, 2002). Because these covariates are strongly related to students' academic success, they were kept in the model regardless of their significance levels. After the set of student-level variables were added, the *City Connects* dummy variable was also added at level-2 to test its significance predicting each of the level-1 slope coefficients. When the *City Connects*' dummy variable was not a significant predictor of the

student-level variables, it was removed from the model. Finally, random effects for the intercepts and slopes were tested and allowed to vary across schools if they were statistically significant.

After several iterations of the model as described above the final model can be expressed as the following:

Level-1 Model:

The Level 1 model equations are as follows:

$$\text{Prob}(\text{EVENT}_{ijk} = 1 \mid \boldsymbol{\beta}_k) = h_{ijk}$$

$$\log[h_{ijk}/(1 - h_{ijk})] = \eta_{ijk}$$

$$\eta_{ijk} = \alpha_{1k} * \text{Grade1}_{ijk} + \alpha_{2k} * \text{Grade2}_{ijk} + \alpha_{3k} * \text{Grade3}_{ijk} + \alpha_{4k} * \text{Grade4}_{ijk} + \alpha_{5k} * \text{Grade5}_{ijk} + \beta_1(Z_{1ijk} - \overline{Z_{1..}}) + \beta_P(Z_{PIJK} - \overline{Z_{P..}})$$

Level-2 Model:

The Level 2 model equations are as follows:

$$\alpha_{1k} = \gamma_{10} + \gamma_{11}CCNX_k$$

$$\alpha_{2k} = \gamma_{20} + \gamma_{21}CCNX_k + u_{2k}$$

$$\alpha_{3k} = \gamma_{30} + \gamma_{31}CCNX_k + u_{3k}$$

$$\alpha_{4k} = \gamma_{40} + \gamma_{41}CCNX_k + u_{4k}$$

$$\alpha_{5k} = \gamma_{50} + \gamma_{51}CCNX_k + u_{5k}$$

$$\beta_{1k} = \gamma_{60}$$

...

$$\beta_{3k} = \gamma_{80} + u_{8k}$$

$$\beta_{4k} = \gamma_{90} + u_{9k}$$

...

$$\beta_{18k} = \gamma_{230}$$

The HLM 7 software does not produce deviance statistics for Bernoulli models when the restricted penalized quasi-likelihood (PQL) method is used for the estimating coefficients. However, it provides the option to use the Laplace estimation to produce the deviance statistic and compare nested models. Thus, while the estimates presented in the tables are from the restricted PQL method, from Model 2A onwards the models were also run using Laplace estimation to compare models and examine whether addition of new variables was justified (Raudenbush, Bryk, Cheong, Congdon, & Toit, 2011). The model fit statistics suggested that addition of each set of student-level variables improved the model fit at each step, with the exception of Model 2C, where the *City Connects* dummy variable was added to predict each of the discrete-time variables. Although this could be because the *City Connects* dummy variable was not statistically significant in predicting some of the discrete-time variables, it was still retained in the model to examine the deviation of *City Connects* from the comparison schools.

Among the models presented, Model 2F was the final version. Model 2F's discrete-time dummy variable coefficients suggest that the average odds of exiting LEP status increased over time for comparison schools, after controlling for student-level variables. The effect of the *City Connects* intervention was captured by the *City Connects* dummy variable. This variable was used in predicting the coefficients of each of the discrete-time variable at the student-level. The magnitude and direction of the *City Connects* dummy indicates on average the deviation of the *City Connects* schools' odds of exiting LEP status from the odds of comparison schools'. In Model 2F, the estimates of the CCNX dummy for each grade are always in the positive direction, indicating that average difference in the odds of exiting LEP status for *City Connects* schools

was higher than comparison schools, after accounting for all student-level covariates, which were centered on their grand-mean. While estimates for the *City Connects* dummy variable were always in the positive direction, two of these estimates, for Grade 2 and Grade 3, were also statistically significant, suggesting that the average difference in odds ratios that exists between the two groups was statistically significant.

The estimated mean odds ratio for the CCNX dummy for Grade 2, $e^{(\beta_{CCNX})}$, was calculated as 1.97 ($\gamma_{21}=0.68$), indicating that the estimated mean odds of exiting LEP status was nearly two times higher for *City Connects* schools than for comparison schools, with all else in the model controlled for. Similarly, for Grade 3, the estimated mean odds ratio was 1.85 ($\gamma_{31}=0.62$), suggesting that the mean odds of exiting LEP was nearly 1.85 times higher for *City Connects* schools than for comparison schools, after all the model covariates were controlled for.

Finally, similar to the one-level models, this model also suggested that girls were 23% (as indicated by 1 divided by boys' odds ratio of 0.81, $p < .01$) more likely than boys to exit LEP status. Similarly, students who did not receive special education services or received regular education only with modifications were 2.3 times more likely to exit LEP than students in SPED 2 (as indicated by 1 divided by SPED 2's odds ratio of 0.43 ($p < .01$)) and 3.7 times more likely to exit LEP than students in SPED 3 (as indicated by 1 divided by SPED 3's odds ratio of 0.27 ($p < .01$)). Moreover, Asian students were twice ($p < .01$) more likely to exit LEP status than their White peers. Finally, students with higher levels of report card scores in math were 24% ($p < .01$), reading 40% ($p < .05$), writing 13% ($p < .05$), and work habits were 15% ($p < .05$) more likely to exit LEP status compared to their peers who scored at the lower levels in these report card scores. These results are, in general, similar with the results of one-level analyses in regard to the variables observed as significant and the direction of the coefficients.

Table 4-12: *Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Big Analytic Sample, ATT Weighted.*

Fixed Effects	Model 2A (Unconditional)				Model 2B			
	Coef.	S.E.	OR	p-value	Coef.	S.E.	OR	p-value
Model for school means								
Intercept	-2.52	0.05	0.08	<0.001				
Grade 1, γ_{10}					-5.47	0.17	0.00	<0.001
Grade 2, γ_{20}					-3.05	0.12	0.05	<0.001
Grade 3, γ_{30}					-2.36	0.12	0.09	<0.001
Grade 4, γ_{40}					-1.57	0.09	0.21	<0.001
Grade 5, γ_{50}					-1.27	0.08	0.28	<0.001
Random Effects	Variance	df	X^2	p-value	Variance	df	X^2	p-value
Intercept	0.13	167	807.1	<0.001				
Grade 2, u_{2k}					0.45	137	450.51	<0.001
Grade 3, u_{3k}					0.68	137	533.36	<0.001
Grade 4, u_{4k}					0.31	137	627.48	<0.001
Grade 5, u_{5k}					0.18	137	284.67	<0.001
Random level-1 coef.	Reliability estimate				Reliability estimate			
Intercept	0.535							
Grade 2					0.498			
Grade 3					0.616			
Grade 4					0.548			
Grade 5					0.417			

Table 4-13: *Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Big Analytic Sample, ATT Weighted.*

	Model 2C				Model 2D			
	Coef.	S.E.	OR	p-value	Coef.	S.E.	OR	p-value
Model for school means								
Grade 1, γ_{10}	-5.61	0.23	0.00	<0.001	-5.84	0.24	0.00	<0.001
CCNX, γ_{11}	0.27	0.33	1.30	0.422	0.44	0.35	1.55	0.205
Grade 2, γ_{20}	-3.37	0.10	0.03	<0.001	-3.54	0.11	0.03	<0.001
CCNX, γ_{21}	0.51	0.19	1.66	0.008	0.64	0.19	1.90	0.001
Grade 3, γ_{30}	-2.51	0.12	0.08	<0.001	-2.76	0.14	0.06	<0.001
CCNX, γ_{31}	0.23	0.20	1.26	0.257	0.49	0.21	1.64	0.018
Grade 4, γ_{40}	-1.57	0.12	0.21	<0.001	-1.68	0.10	0.19	<0.001
CCNX, γ_{41}	-0.03	0.18	0.97	0.873	0.11	0.16	1.11	0.506
Grade 5, γ_{50}	-1.32	0.08	0.27	<0.001	-1.42	0.08	0.24	<0.001
CCNX, γ_{51}	0.07	0.15	1.07	0.660	0.25	0.14	1.28	0.080
Model for slopes								
Male, γ_{60}					-0.35	0.04	0.71	<0.001
Black, γ_{70}					-0.26	0.15	0.77	0.089
Asian, γ_{80}					0.82	0.17	2.28	<0.001
Hispanic, γ_{90}					0.01	0.13	1.01	0.969
Other, γ_{100}					0.13	0.23	1.14	0.559
Random Effects								
	Variance	df	X^2	p-value	Variance	df	X^2	p-value
Grade 2, u_{2k}	0.40	136	427.05	<0.001	0.41	121	262.88	<0.001
Grade 3, u_{3k}	0.66	136	537.32	<0.001	0.77	121	379.41	<0.001
Grade 4, u_{4k}	0.32	136	637.04	<0.001	0.23	121	295.79	<0.001
Grade 5, u_{5k}	0.18	136	289.35	<0.001	0.18	121	174.48	<0.001
Asian, u_{8k}					0.29	122	238.05	<0.001
Hispanic, u_{9k}					0.11	122	225.91	<0.001

Table 4-13 (Continued): *Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Big Analytic Sample, ATT Weighted.*

Random level-1 coef.	Reliability estimate			Reliability estimate		
Grade 2	0.473			0.420		
Grade 3	0.611			0.541		
Grade 4	0.550			0.416		
Grade 5	0.414			0.346		
Asian				0.271		
Hispanic				0.258		
	χ^2	df	p-value	χ^2	df	p-value
Model Fit Statistics	7.547	5	0.182	334.3	16	<0.001

Table 4-14: *Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Big Analytic Sample, ATT Weighted.*

	Model 2E				Model 2F			
	Coef.	S.E.	OR	p-value	Coef.	S.E.	OR	p-value
Model for school means								
Grade 1, γ_{10}	-6.00	0.24	0.00	<0.001	-6.25	0.25	0.00	<0.001
CCNX, γ_{11}	0.44	0.35	1.55	0.210	0.46	0.34	1.59	0.178
Grade 2, γ_{20}	-3.66	0.11	0.03	<0.001	-3.86	0.12	0.02	<0.001
CCNX, γ_{21}	0.65	0.19	1.91	0.001	0.68	0.20	1.97	0.001
Grade 3, γ_{30}	-2.85	0.14	0.06	<0.001	-3.03	0.14	0.05	<0.001
CCNX, γ_{31}	0.51	0.21	1.67	0.015	0.62	0.22	1.85	0.005
Grade 4, γ_{40}	-1.74	0.10	0.18	<0.001	-1.85	0.11	0.16	<0.001
CCNX, γ_{41}	0.13	0.16	1.14	0.432	0.27	0.17	1.31	0.124
Grade 5, γ_{50}	-1.45	0.08	0.23	<0.001	-1.41	0.09	0.24	<0.001
CCNX, γ_{51}	0.28	0.14	1.32	0.051	0.30	0.16	1.35	0.058
Model for slopes								
Male, γ_{60}	-0.29	0.04	0.75	<0.001	-0.21	0.05	0.81	<0.001
Black, γ_{70}	-0.22	0.14	0.80	0.109	-0.10	0.15	0.91	0.513
Asian, γ_{80}	0.76	0.15	2.13	<0.001	0.72	0.16	2.05	<0.001
Hispanic, γ_{90}	0.09	0.12	1.10	0.421	0.27	0.13	1.31	0.035
Other, γ_{100}	0.15	0.22	1.16	0.488	0.31	0.23	1.37	0.178
SPED2, γ_{110}	-1.11	0.11	0.33	<0.001	-0.84	0.11	0.43	<0.001
SPED3, γ_{120}	-1.97	0.19	0.14	<0.001	-1.32	0.20	0.27	<0.001
Reduced-Price Lunch, γ_{130}	-0.04	0.15	0.96	0.760	0.11	0.16	1.12	0.488
Free-Price Lunch, γ_{140}	-0.52	0.16	0.60	<0.001	-0.18	0.16	0.83	0.252
Foreign-Born, γ_{150}	0.05	0.07	1.05	0.473	0.23	0.08	1.26	0.002
Mobile before Gr1, γ_{160}	-0.30	0.34	0.74	0.384	-0.33	0.38	0.72	0.391
Retained before Gr1, γ_{170}	-0.04	0.19	0.96	0.817	-0.01	0.19	0.99	0.937

Table 4-14 (Continued): *Estimated City Connects Treatment Effects in Elementary School on Exiting LEP Status with Robust Standard Errors, Big Analytic Sample, ATT Weighted.*

Z_Math, γ_{180}					0.21	0.03	1.24	<0.001
Z_ELA, γ_{190}					0.34	0.07	1.40	<0.001
Z_Writing, γ_{200}					0.13	0.06	1.13	0.044
Z_Effort, γ_{210}					0.01	0.04	1.01	0.773
Z_Work Habits, γ_{220}					0.14	0.05	1.15	0.004
Z_Behavior, γ_{230}					0.00	0.04	1.00	0.971
Random Effects	Variance	df	X^2	p-value	Variance	df	X^2	p-value
Grade 2, u_{2k}	0.41	121	275.19	<0.001	0.45	121	282.00	<0.001
Grade 3, u_{3k}	0.76	121	372.93	<0.001	0.88	121	355.74	<0.001
Grade 4, u_{4k}	0.25	121	304.90	<0.001	0.31	121	307.49	<0.001
Grade 5, u_{5k}	0.18	121	178.84	<0.001	0.26	121	182.57	<0.001
Asian, u_{8k}	0.28	122	231.40	<0.001	0.33	122	229.68	<0.001
Hispanic, u_{9k}	0.14	122	292.83	<0.001	0.17	122	271.64	<0.001
Random level-1 coef.	Reliability estimate				Reliability estimate			
Grade 2	0.418				0.425			
Grade 3	0.539				0.547			
Grade 4	0.420				0.441			
Grade 5	0.346				0.385			
Asian	0.264				0.277			
Hispanic	0.287				0.310			
Model Fit Statistics	X^2	df	p-value		X^2	df	p-value	
	105827.82	27	<0.001		957.9	6	<0.001	

Figures 4-5 and 4-6 present the fitted hazard and survival probabilities for the big analytic sample, respectively. In each graph, the blue lines show the estimates for the *City Connects* and the red lines comparison schools. The hazard and survival probabilities were estimated using the same procedures as described for sub-samples 1 and 2.

Figure 4-5 reveals that the pattern of hazard was similar for the *City Connects* and comparison group. For the big analytic sample, the conditional probability of exiting LEP status increased over time and it was the highest in Grade 5, given that students had not exited the LEP status previously. Examining the difference between the *City Connects* and comparison schools, note that there was a vertical separation between the two groups. Specifically, this difference was the largest in Grade 5. Lastly, figure 4-5 suggests that the probability of exiting LEP for students in the *City Connects* schools was nearly two times greater than that of comparison school students in Grade 2 and nearly 1.8 times more than that of comparison school students in Grade 3.

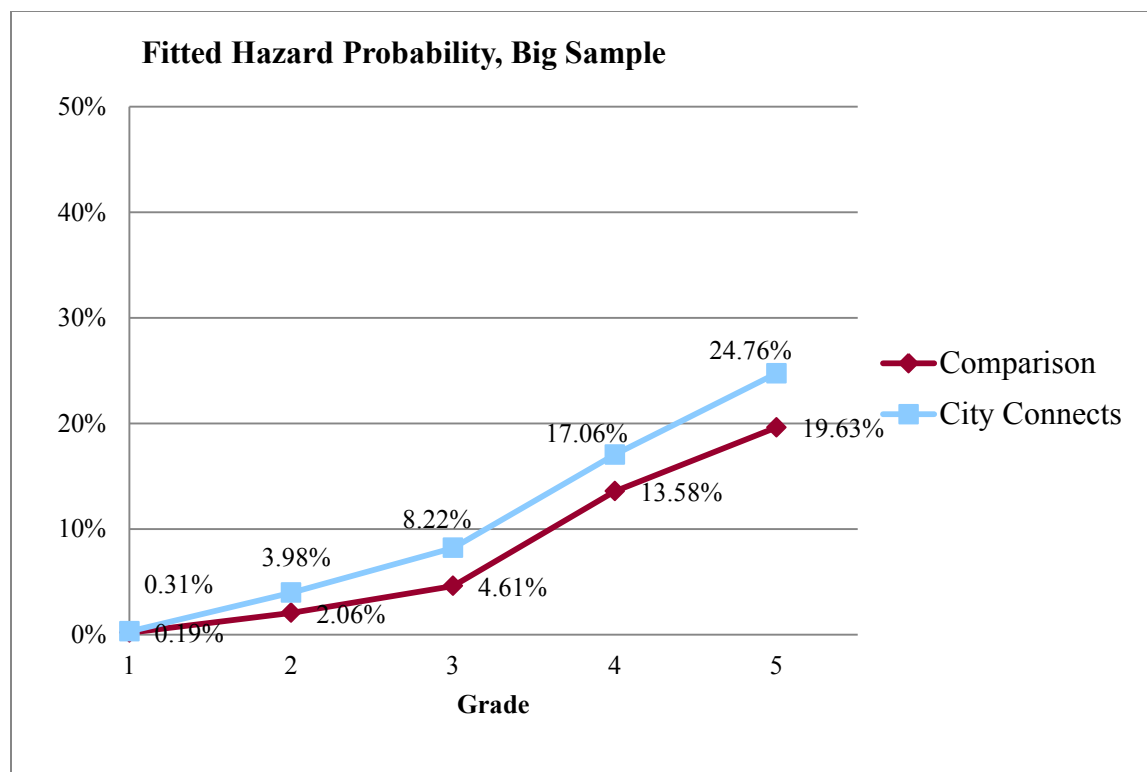


Figure 4-5: Fitted hazard probability of exiting LEP status conditional on student characteristics, Big Analytic Sample.

Figure 4-6 provides a graphical summary of the survival probabilities as students in the big analytic sample progressed from one grade to the next for the two groups. Note that a survival trend always goes downward since a participant who experienced the event can never return to the risk set. As indicated by the steeper downward slope for the *City Connects* group, the percent of students exiting LEP status was in general more for this group than for the comparison group. Figure 4-6 reveals that, by the end of the fifth grade, there were nearly 9.93% more students who were still in the LEP status in the comparison schools than in the *City Connects* schools.

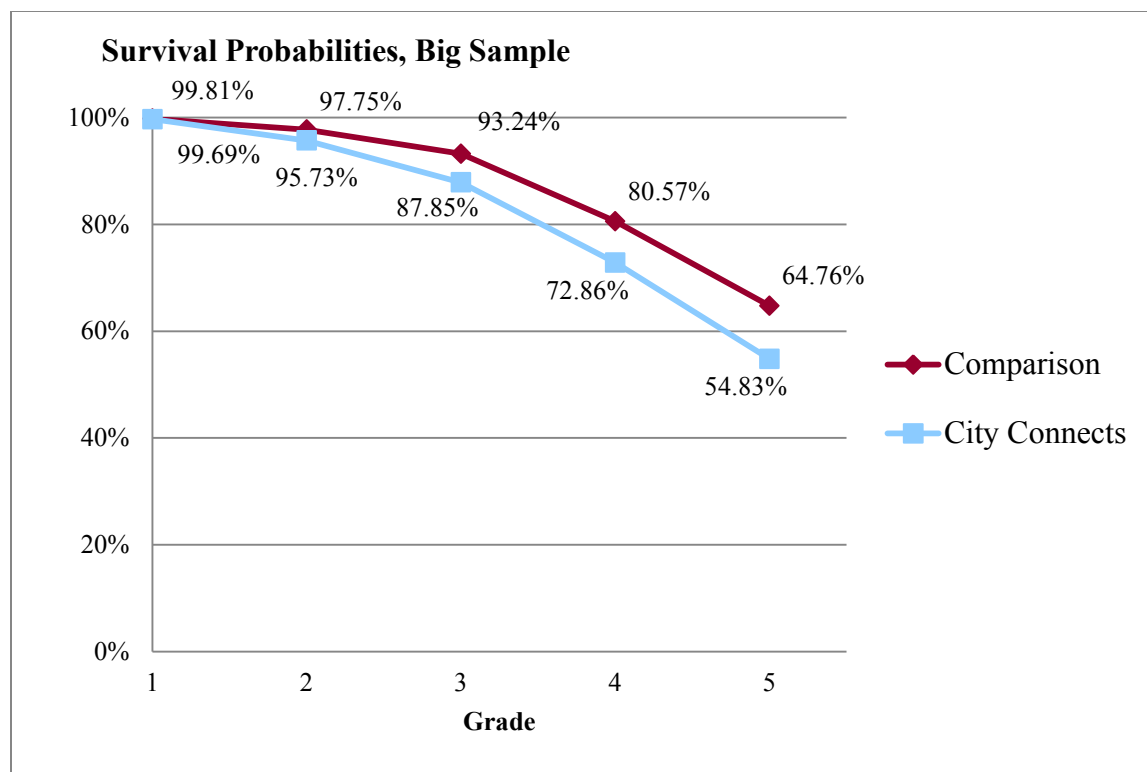


Figure 4-6: Survival probability conditional on student characteristics, Big Analytic Sample.

The cumulative hazard rate for each group was estimated simply by subtracting the percent of students that were still in the LEP status in Grade 5 in Figure 4-6 from 100%. This yielded values of 45.17% for the students in *City Connects* and 35.24% for the students in comparison schools. In other words, nearly 10% more students in *City Connects* schools had exited LEP status by the end of fifth grade than students in comparison schools.

4.4 Research Question Three

The third research question concerns the median time to exit LEP status by the *City Connects* and the comparison groups. To answer this question, the plot of survival probabilities for Model 2F for the big analytic sample was used, where all the covariate values were held equal to the overall average of the schools in the big analytic sample. In this plot, the point in time corresponding to the .5 survival probability indicates the median time to exit LEP status for each group, i.e., the time by which half of the students had exited LEP status.

Figure 4-7 repeats Figure 4-6 but also displays the extrapolated parabolas that were fitted to the survival probabilities for each group for fitted Model 2F. Solving the *City Connects*' equation with y equal .5 yields x equal 5.2, suggesting that about half of the *City Connects*' students will have exited LEP status by the first quarter of sixth grade. Likewise, solving the comparison schools' equation of the parabola with y equal .5 yields x equal 5.7, suggesting that about half of students in this group will have exited LEP status by the third quarter of sixth grade (BPS issues four report cards per year for middle schools for Grades 6 through 8 (including Grades 6-8 in K-8 schools) (BPS, 2015)). Accordingly, while about half of the LEP students that graduated from *City Connects* elementary schools could start enrolling in mainstream classes no later than the second quarter of the sixth grade, the same proportion of LEP students in the comparison group would exit LEP status no later than the third quarter of the sixth grade and could either begin enrolling in mainstream classes at the end of sixth grade or beginning of the seventh grade. In other words, for a typical student in a *City Connects* school, this may translate into a gain of at least one half of a year in grade in mainstream classes. However, note that this study duration only encompasses the elementary grades. Thus, the plausible changes in school environment and policies from elementary to middle grades may impact the time to exiting LEP status in ways that this study was not able to measure. As a result, the median time to the event that goes beyond five years may be biased to some extent.

If the same procedure is repeated using the survival probabilities from research question 1 for the big analytic sample, the median life time is 4.9 years for the *City Connects* group while it is 5.5 years for the comparison group. This suggests that with no modelling and conditioning on any covariates, basic life-table analysis suggests that median time to exit LEP status is nearly one quarter shorter for each group compared to the Model 2F. Also, note that while the plausible

gain of at least one half of year in grade in mainstream classes remains the same, because median time for the comparison group still goes beyond five years, bias in this estimate still remains as a concern.

In order to eliminate the concern resulting from to median time to the event going beyond five years, the above procedure was repeated with the point in time corresponding to the .67 survival probability for each group, i.e., the time by which one third of the students had exited LEP status. For the *City Connects*' extrapolated parabola, y equal .67 yields x equal 4.4, suggesting that roughly one third of the *City Connects*' students exited LEP status by the first half of fifth grade. Similarly, solving the comparison schools' extrapolated parabola with y equal .67 yields x equal 4.9, suggesting that roughly one third of students in this group have exited LEP status by the end of fifth grade (BPS issues three report cards per year for elementary schools for grades K-5 (including K-5 in K-8 schools) (BPS, 2015)). In other words, while about one third of the LEP students in *City Connects* elementary schools could start enrolling in mainstream classes no later than the beginning of the third semester of the fifth grade, the same proportion of LEP students in the comparison group could do so no later than the beginning of the sixth grade. If the same procedure is repeated using the survival probabilities from research question 1 for the big analytic sample, the life time that corresponds to .67 is 4 years for the *City Connects* group while it is 4.5 years for the comparison group. This suggests that with no modelling and conditioning on any covariates, basic life-table analysis indicates that time to exit LEP status is nearly two quarters shorter for each group. But, the gain of at least one half of year in grade in mainstream classes remains the same.

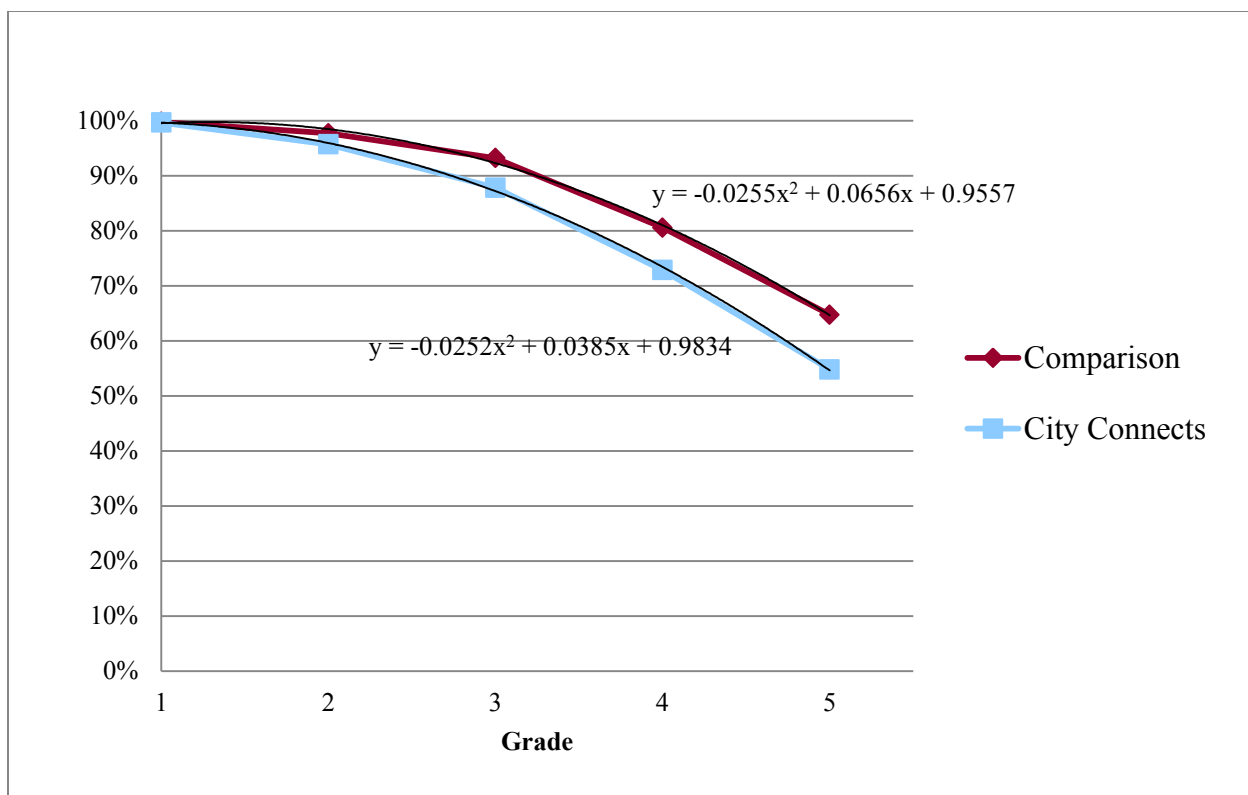


Figure 4-7: Fitted trend lines for survival probabilities, Big Analytic Sample.

4.5 Research Question Four

The last research question asks whether the estimated treatment effects are robust to the presence of unobserved selection bias. To answer this question a sensitivity analysis was conducted using Model 2F from research question two.

In Chapter 3, it was hypothesized that parental involvement constituted an important but unobserved variable, designated U , related to both treatment assignment R and outcome Y . In order to introduce selection bias to this study, two assumptions were made about the unobserved variable U . These were: 1) U was probabilistically related to treatment assignment R . For example, parents who were highly involved with their children's education were more likely to enroll them in a *City Connects* school than in a comparison school. In other words, the conditional probability of U given R needed to satisfy $\mu_{11} > \mu_{10}$. 2) The unobserved variable U

is strongly positively related to the outcome of exiting LEP status. That is, the regression coefficient for U (β_U) will be fixed as a positive value when the simulated variable U is introduced as a student level covariate in the HLM model.

To carry out the sensitivity analysis, first, the unobserved variable U was simulated ten times using the ten pairs of conditional probabilities of U given R . These ten pairs were previously specified in Table 3-6. Note that the data at hand is longitudinal in nature. Thus, the unobserved variable U was first simulated for each unique student record in the data set. The simulated values were then copied to the other records with the same student id. Next, examining Model 2F, the value β_U was set to 0.8, a value greater than the magnitude of student level covariates similar to parental involvement in Model 2F. In this model, the highest two coefficients belonged to Reading Report Card score (Z_ELA) and foreign-born status, 0.34 and 0.23 in log odds, respectively. A value stronger (i.e., larger) than the values representing both prior achievement and foreign-born status was chosen. To adjust the binary outcome of exiting LEP status by the simulated U , the simulated U variables were multiplied by 0.8 and then were included in Model 2F one at a time.

Table 4-15 presents the estimated treatment effects of the *City Connects* intervention for each discrete-time variable when the outcome model was adjusted for the simulated U values and their pre-determined relationship of 0.8 to the outcome variable. This analysis was carried once per each simulated U . Table 4-15 also displays the actual treatment effects from Model 2F, their corresponding standard errors, and the 95% and 90% confidence intervals, which are at the bottom of the table.

Table 4-15: *The Estimated Treatment Effects of the City Connects Intervention after Model 2F Was Adjusted for Simulated U Values and the Fixed Regression Coefficient of 0.8.*

U	μ_{10}	μ_{11}	γ_{11}	γ_{21}	γ_{31}	γ_{41}	γ_{51}				
u_1	0.20	0.35	0.47	0.69	0.63	0.28	0.31				
u_2	0.20	0.50	0.47	0.68	0.62	0.28	0.30				
u_3	0.20	0.65	0.47	0.68	0.63	0.28	0.30				
u_4	0.20	0.80	0.49	0.71	0.65	0.30	0.33				
u_5	0.35	0.50	0.46	0.68	0.62	0.27	0.30				
u_6	0.35	0.65	0.46	0.68	0.62	0.27	0.30				
u_7	0.35	0.80	0.44	0.66	0.60	0.25	0.28				
u_8	0.50	0.65	0.46	0.67	0.61	0.26	0.29				
u_9	0.50	0.80	0.46	0.67	0.61	0.26	0.29				
u_{10}	0.65	0.80	0.46	0.68	0.62	0.27	0.30				
			γ_{11}	γ_{21}	γ_{31}	γ_{41}	γ_{51}				
			0.46	0.68	0.62	0.27	0.30				
	S.E.		0.34	0.20	0.22	0.17	0.16				
	Two Sided 95% CI	-0.21	1.13	0.28	1.07	0.19	1.05	-0.07	0.61	-0.01	0.60
	Two Sided 90% CI	-0.10	1.03	0.34	1.01	0.26	0.98	-0.02	0.56	0.04	0.55

Table 4-15 reveals that estimated treatment effects either slightly decreased or increased when a coefficient of 0.8 was employed. Also, the estimated coefficients of γ_{11} through γ_{51} for each pair of simulated U fall within the 90% and 95% confidence intervals of the original ones. Thus, the estimated treatment effects can be assumed to be reasonably robust to the presence of the type of hidden bias investigated in this study. However, note that the sensitivity analysis presented here has its limitations. It was carried out with only ten simulated U s based on the ten pairs of conditional probabilities of U given R . For a sensitivity analysis to be considered more conclusive, a greater number of simulations are, in general, recommended. Also, the longitudinal nature of the data may require more investigation in regard to appropriateness of the procedures used. Thus, the sensitivity analysis conducted here should be considered exploratory in nature. Nonetheless, the results provide a glimpse into the extent to which the estimates are robust to some form of a hidden bias.

Chapter 5. Conclusions

The analyses presented in this study were designed to estimate the effect of participation in the *City Connects* intervention has on the likelihood of a student's exiting LEP status during the elementary grades. This chapter discusses the study's results, limitations, recommendations for future research, and presents policy implications.

5.1 Summary of Findings

To evaluate *City Connects* treatment effects on the likelihood of exiting LEP status during elementary grades, students' longitudinal records were used to conduct parallel analyses that included three samples. The hypotheses underlying this study were that students enrolled in *City Connects* schools tend to thrive and achieve more than their counterparts in non-*City Connects* schools because the *City Connects* intervention fosters students' academic, social-emotional, and physical well-being. Specifically, the intervention provides a student support system that evaluates each student individually and then meets that student's specific needs by fostering his/her strengths through a tailored set of community-based prevention, intervention, and enrichment services. Thus, as one measure of LEP students' academic achievement, exposure to the *City Connects* intervention would be expected to contribute to their likelihood of exiting LEP status earlier than otherwise. A summary of the empirical findings in support of these arguments is presented below.

5.1.1 Research question 1.

In order to investigate whether the data at hand suggested an association between likelihood of exiting of LEP status and attending a *City Connects* school, this study employed life table analysis. The life tables generated for the three samples reported detailed information on the number of students who were LEP at the beginning of a grade, the number who were

censored at the end of that grade, and the number who exited LEP status during that grade. With these numbers, the proportions of students who remained in LEP status at the end of a grade and who exited LEP status by the end of that grade were calculated. For all three samples, these tables suggest that LEP students enrolled in *City Connects* schools were more likely to exit LEP status than students enrolled in comparison schools.

While life table analysis provided detailed information on the proportions of students exiting LEP status at each elementary grade, it was not capable of testing whether the survival distributions differed overall between the two groups. Thus, a log-rank test was used to compare the Kaplan-Meier curves of the *City Connects* and comparison groups for the three samples, and the results suggested that the survival distributions were statistically not equal. In other words, observing that the two groups differed in their survival provided the initial evidence prompting further examination of the event of *exiting LEP status* with more advanced modeling to better understand the effect of the *City Connects* intervention on this event.

5.1.2 Research question 2.

The findings related to the analyses for research question 2, which employed more advanced methods and incorporated clustering based on school- and student-level covariates, further confirmed the findings for research question 1. One challenge in conducting research concerning the effects of participation in *City Connects* were the pros and cons of performing the analysis accounting for the true level at which treatment assignment occurred. Specifically, participation in and adoption of the *City Connects* intervention occurred at the school level rather than at student level.

Since its first implementation in the early 2000s, the number of elementary schools using *City Connects* throughout the BPS has gradually increased. The pattern of *City Connects*'s

expansion over time, however, limits the capability to perform analysis in some ways. For example, a larger number of schools would be needed in order to perform a two-level analysis that would properly account for school-level participation. Accordingly, obtaining a sufficiently large sample for analysis, including as many *City Connects* schools as possible, was necessary. However, a drawback of this approach included the different years in which the schools adopted the *City Connects* intervention. Employing differing starting years made establishing a school-level baseline equivalence problematic as choosing which baseline years should be taken into account for the comparison schools was difficult. As a consequence, previous research on *City Connects* has used a two-level analysis and proceeded by establishing a student-level baseline equivalence rather than a school-level one. Due to the same concern, this study employed a similar approach, included all schools whose data were suitable for analysis in the big analytic sample and then carried out a two-level discrete-time event history analysis with balancing weights at the student-level.

To account for the true treatment assignment level, this study used a school-level matching approach by focusing on a sub-sample of *City Connects* schools that started the intervention in the same year. Following this approach, this study established sub-samples 1 and 2. Sub-sample 1 included 18 schools in total: six *City Connects* schools that began using the intervention for the first time in 2001 and 12 matched comparison schools. Sub-sample 2 included 12 schools in total: four *City Connects* schools that started using the intervention for the first time in 2007 and eight matched-comparison schools. The matched-comparison schools were identified using an optimal matching method and the schools' last three years of pre-intervention data obtained from NCES CCD. One-level analyses were performed on these sub-samples, with STATA's *vce* (cluster *clustvar*) option, which accounted for school clustering.

With the school-matching approach, while this study was able to establish baseline equivalence at the school level, it was also able to incorporate student-level balancing weights. That is, once the samples were formed using optimal matching, within each sample, student-level baseline equivalence was then established. The only drawback with this approach was that sample sizes were smaller for sub-samples 1 and 2 than for the big analytic sample. However, observing similar results on all three of the samples served to lend additional credence to the empirical evidence observed for the effectiveness of the *City Connects* intervention.

On sub-samples 1 and 2, the *City Connects* effects were estimated through a one-level discrete-time event history model, with STATA's *vce* (cluster *clustvar*) option addressing clustering and providing robust estimates of standard errors. ATT weights were applied to reduce explicit bias, and key student characteristics were included as covariates. The analysis indicated that the *City Connects* intervention had a substantial impact on students' likelihood of exiting LEP status. Specifically, the magnitudes of this effect were 0.33 and 0.38 in log odds for sub-samples 1 and 2, respectively. These effects were statistically significant and translated into odds ratios of 1.4 and 1.5, respectively. In other words, the odds of *City Connects* students exiting LEP status was, on average, 40% higher than those attending the matched-comparison schools in sub-sample 1 in any grade. Similarly, on average, the odds of *City Connects* exiting LEP status was 50% higher than that of students in the matched-comparison schools in sub-sample 2 in any grade. In sub-sample 1, of all LEP students in the *City Connects* schools, 57% exited LEP status by the end of the fifth grade whereas of all LEP students in the matched-comparison schools, 46% exited LEP status by the end of the fifth grade. For sub-sample 2, these rates were 43% and 32% for students in the *City Connects* and matched-comparison schools, respectively.

The two-level discrete-time event history model applied to the big analytic sample yielded similar results. With this model, after adjustment for key student-level covariates and school-level random effects, the effect of *City Connects* participation on exiting LEP status was estimated for each elementary grade. Figure 4-5 displays the conditional likelihood of exiting LEP status in each grade for the *City Connects* and comparison schools based on Model 2F. As can be seen in the figure, as students of both groups progressed through the elementary grades, their likelihood of exiting LEP status in a particular grade increased, with the fifth grade being the most likely time at which this would occur, given that they had not exited LEP status prior to entering that grade. In answer to the question as to whether attending a school participating in the *City Connects* intervention influenced the likelihood of a student exiting LEP status, the difference between the conditional likelihoods in the plots for the two groups in Figure 4-5 provides an answer. Comparison of the plots reveals that LEP students enrolled in *City Connects* schools were significantly more likely to exit LEP status in Grades 2 and 3 than LEP students attending comparison schools. The odds ratios associated with Grades 2 and 3 were 1.97 and 1.85, respectively. Thus, the mean odds of exiting LEP status in *City Connects* schools was nearly double that of comparison schools in Grade 2 and 3, after accounting for all the student level variables. Lastly, with respect to the cumulative hazard rate, of all LEP students enrolled in *City Connects* schools, 45% had exited LEP status by the end of Grade 5, whereas only 35% of those enrolled in comparison schools had.

A socio-ecological framework provides a useful approach for explaining the treatment effects observed for the *City Connects* intervention. According to Bronfenbrenner's (2009) socio-ecological framework, microsystems encompass the environments with which students have immediate contact, including their schools, families, and neighborhoods. At this level, the quality

of students' bi-directional relationships with these entities helps shape their immediate learning environments (Bronfenbrenner, 2009). Mesosystems, on the other hand, are those systems that connect different microsystems and enable communication between them (Bronfenbrenner, 2009). For example, the connection between a student's parents and teachers and between the student's school and neighborhood can each be considered a mesosystem. In this context, the *City Connects* intervention can be viewed as a mesosystem that connects several microsystems surrounding students. The *City Connects* intervention fills this role in a systematic way by providing prevention, intervention, and enrichment programs delivered through a network of interrelated partnerships between school, family, community, and university. The positive findings observed in this study support the hypothesis that an intervention incorporating multiple ecosystems that affect students' lives, such as *City Connects*, is an effective intervention.

5.1.3 Research question 3.

In research question 2 and 3, median lifetimes were estimated for each sample as a summary statistic using the survival probabilities based on the final fitted models (see Figure 4-7). Table 5-1 summarizes these median lifetimes. A median lifetime is the point in time by which half of a sample has experienced an event and half has not. The lifetime by which a third of the sample has experienced the event was also estimated, but only for the big-sample.

Table 5-1: *Median Lifetimes based on final fitted models for each sample.*

	Time to event when	
	survival probability=0.5	survival probability=0.67
Model 2F- Big Sample		
City Connects	5.2	4.4
Comparison	5.7	4.9
Life Table -Big Sample		
City Connects	4.9	4
Comparison	5.5	4.5
Model 1E- Sub Sample 1		
City Connects	4.7	
Matched-Comparison	5.1	
Model 1E- Sub Sample 2		
City Connects	5.2	
Matched-Comparison	5.9	

Note 1: BPS issues three report cards per year for elementary schools for grades K-5 (including K-5 in K-8 schools) (BPS, 2015)

Note 2: BPS issues four report cards per year for middle schools for Grades 6 through 8 (including Grades 6-8 in K-8 schools) (BPS, 2015).

Findings indicated that half of the *City Connects*' students had exited LEP status either at or before the end of Grade 5 or at or before the first quarter of Grade 6. For the comparison group, half of the students had exited LEP status at or before the second half of Grade 6, except for sub-sample 1, in which half of the comparison students had exited LEP status at or before the first quarter of Grade 6. In other words, for a typical student in a *City Connects* school, these median lifetimes may translate into a gain of at least one half of a year in grade in mainstream classes. However, as previously mentioned in Chapter 4, note that this study duration only encompasses the elementary grades. Thus, the plausible changes in school environment and policies from elementary to middle grades may impact the time to exiting LEP status in ways that this study was not able to measure. As a result, the median time to the event that goes beyond five years may be biased to some extent.

In order to remedy this concern, a similar lifetime estimation was done for the first tierce based on Model 2F, where all the covariate values were held equal to the overall average of the schools in the big analytic sample. As it was expected, the lifetimes for when the third of the sample had experienced the event were shorter, nearly three quarters, than the median lifetimes. However, the difference between the *City Connects* and the comparison groups remained the same; that is, a gain of at least one half of a year in grade in mainstream classes.

The median lifetimes observed in this study are consistent with the limited literature on this topic; that it takes approximately four to seven years for a typical LEP student to be reclassified as English proficient (Grissom, 2004; Hakuta et al., 2000; Mavrogordato, 2012; Parrish et al., 2006; Slama, 2012; Thompson, 2012). In addition to the difference in median lifetimes, considering the 10% difference in cumulative hazard rates by the end of Grade 5, these findings, overall, are positive. They suggest that the student support systems that *City Connects* put in place could translate into more time in mainstream classrooms for more students. These differences between the two groups could be crucial, because *City Connects* students' entering mainstream classes earlier on in their school careers could translate into important academic and non-academic gains, such as increased self-confidence and better scores on academic assessments.

5.1.4 Research question 4.

The research on the *City Connects* intervention was conducted via a quasi-experimental design since schools were not randomly selected to adopt the intervention. This lack of random assignment inevitably raises concerns regarding selection bias since schools and students in the treatment group could systematically differ from those in the comparison group (Schneider, Carnoy, Kilpatrick, Schmidt, & Shavelson, 2007). In the context of the *City Connects*

intervention, bias translates into possible systematic differences between students attending *City Connects* schools and students attending comparison schools. In the case of explicit/overt bias, statistical controls using observed pre-treatment variables can be included in the models to reduce this bias. For example, such adjustments include establishing baseline equivalences using propensity score weights or propensity score matching. However, in the case of a hidden bias, statistical adjustment to analyses is more complicated since the variable causing selection bias is unobserved or unmeasured and so is not captured in the data. In this research, if hidden bias existed, then estimated average differences in the likelihood of exiting LEP status between *City Connects* and comparison school students would be biased to some extent, corresponding to over- or under-estimation of treatment effects.

Through a sensitivity analysis using methods proposed by Rosenbaum and Rubin (1983a) and Montgomery et al. (1986), this study explored the degree to which *City Connects*' treatment effects were robust to the presence of unobserved selection bias. Specifically, an unobserved variable 'parental involvement,' designated as U , was hypothesized, and two assumptions were made concerning this unobserved variable. The first assumption was that parents who were highly involved with their children's educations were more likely to enroll them in a *City Connects* school than in a comparison school and that parents who were less involved with their children's education were more likely to enroll them in a comparison school than in a *City Connects* school. The second assumption was that, everything else being equal, the likelihood of a student exiting LEP status was larger when parental involvement was high compared to that of a student with relatively less involved parents. With respect to the first assumption, the unobserved variable U was generated using ten different levels of conditional probabilities of parental involvement given students were in a *City Connects* or a comparison school. With

respect to the second assumption, the regression coefficient of U (i.e., β_U) was set to a positive value. To determine the magnitude of β_U , the magnitude of all available student level covariates similar to parental involvement were examined and then a value greater than the highest positive regression coefficient was assigned to the parental involvement variable U .

The results obtained from sensitivity analyses revealed that the estimated treatment effects associated with *City Connects* participation were either reduced or increased slightly with the inclusion of U in the prediction model. These estimates, however, still fell within the 90% confidence intervals of the original ones, and, thus, estimated treatment effects can be considered to be reasonably robust to the presence of the hidden bias specified for this study.

However, it is important to note that the results presented here were limited to some degree. The sensitivity analysis was carried out with only ten simulated U s based on the ten pairs of conditional probabilities of U given the treatment conditions. While a greater number of simulations would produce more conclusive results, the magnitude of β_U could be varied rather than being fixed at a single value. Thus, the sensitivity analysis conducted as part of this research should be considered exploratory in nature. Although results provide a sense of the robustness of the estimates to some form of hidden bias, its exploratory nature can be considered as a limitation. Thus, future research is suggested which focus solely on sensitivity analysis of the *City Connects* effects.

5.2 Limitations and Future Research

The estimated effects of the *City Connects* intervention on LEP students are of importance for the field of research on ELs, as well as on policy and practice. Given the limited research in this area, in particular, exploring the likelihood of exiting LEP status earlier than otherwise, due to participation in an intervention that addresses out-of-school barriers that affect

learning has wide implications. Despite the importance of this study's subject and the methodologies that it employed to account for the plausible effects of selection bias, this study has seven methodological limitations.

First, the models could have been improved in three ways: 1) by including school-level contextual factors, 2) by including cohort variables, and 3) by implementing a better evaluation approach for the model fit. These three limitations with regard to improving the models are discussed next.

- The designs of the analyses did not allow for controlling for school-level contextual factors. For sub-samples 1 and 2, although school matching was done using the data averaged across three pre-intervention years using NCES CCD school-level data, the discrete event history analyses were performed at the student-level. Thus, school-level contextual factors were not included in these analyses. However, the school-level matching performed for sub-samples 1 and 2 partially compensated for this limitation. For the two-level analysis, with respect to the time-varying aspect of the student-level data, aggregating the school-level data over time for different years became overly complicated. For example, some schools did not have data for some school years. As a result, school-level contextual factors were not included in the two-level analysis for the big analytic sample. If school-level contextual factors had been included in the two-level analysis, their interaction with the *City Connects* indicator could have helped explain some of the variation across *City Connects* schools. For example, when schools are compared with regard to their EL students' countries of origin, it is very common to see that a group of EL students from similar backgrounds (i.e., who were born in the same countries or whose families immigrated from the same countries) attend the same school.

Depending on the characteristics of the community of EL students gathered in one school, the differences between these communities may lead to either better or worse outcomes for the EL students. For example, families in some immigrant communities may support each other in better adapting to school systems, which may help newly arrived families to receive the services that are available to them sooner. For example, undocumented families may learn about their rights sooner in such communities or may be more likely to participate in the federal nutrition assistance program. Thus, having school-level contextual factors could have helped explain not only some of the variation between *City Connects* and comparison schools, but also the variation among *City Connects* schools.

- The cohort variables (i.e., cohort 2001 through 2012) indicating the school year when each cohort's students were in kindergarten could have been included at the student-level for both the one- and two-level models. These variables might have accounted for some of the history effects, as educational policies and schools change over time. Also, the interaction of these variables with the *City Connects* indicator might have explained the changes over time for the *City Connects* intervention. While it is preferable to include these variables in the models, this study could not do so due to the small school sample sizes in the one-level models. Also, to keep the models parsimonious, cohort variables (12 in total) were not included in the two-level models. This was because including the cohort variables and examining their interaction with the *City Connects* indicator would increase the model complexity such that either the sample size might not support the model or the complexity might lead to difficulty in interpreting the results.

- The coefficient of determination (R^2) as a summary measure indicating the goodness-of-fit in linear regression models is easy to interpret as it ranges between 0 and 1 and can be expressed as the proportion of the variance explained by the model with respect to the total variance to be explained. However, the coefficient of determination is not produced in logistic regression as part of the model evaluation statistics. Thus, model evaluation becomes more challenging in logistic regression. While in one-level logistic regression models, the model fit statistics only tests whether the regression coefficients of the newly introduced variables are equal to zero (Wald test), for the two-level models it is not possible to obtain the deviance statistics directly before changing the estimation method from PQL to Laplace estimation. However, over the past decade, an analog to the coefficient of determination has been proposed, called coefficient of discrimination (D), for logistic regression models (Tjur, 2009). The D statistics also ranges between 0 and 1 and simply corresponds to the difference between the means of predicted values of the dependent variable when the outcome is 1 and when the outcome is 0. Basically, once the logistic regression is performed, the predicted probabilities for the outcome variable is stored in the dataset, and then, a t-test is conducted to estimate the difference between the means of the predicted probabilities for the two categories of the outcome. This measure indicates “model’s ability to discriminate between successes and failures” (Tjur, 2009, pg.9). This study was not able to use the coefficient of discrimination in evaluating the model fit as I only became aware of this method during the final stages of writing of this dissertation study.

Second, this study was focused on average treatment effects for the students who entered BPS schools at latest by the start of first grade. The main reason was to isolate the treatment

effects associated with attending school in BPS (i.e., to minimize any confounding that might occur due to students attending schools in other districts before enrolling in BPS schools). The second reason was to avoid left-hand censoring, which is more complicated to address in discrete-time event history analysis. However, it is worth noting that many LEP students enroll in BPS schools after first grade, and so it would be useful to conduct an analysis to examine the effect of *City Connects* on these LEP students, who were excluded from this study.

Third, in this study, a binary indicator of a student's having ever attended a *City Connects* school was used to estimate the treatment effect of the *City Connects* intervention. However, the lengths of time students spent in *City Connects* schools differed, thus affecting the dosage of the *City Connects* intervention that students received. Although prior studies on *City Connects* found a positive association between dosage level and academic success (Walsh et al., 2014), due to the methodological nature of this study, specifying dosage was not possible. Specifically, in the context of this study, the measure of the *City Connects* intervention's effectiveness was how early in their schooling (i.e., the grade and quarter within the grade) students exited LEP status. In other words, students who exited LEP status early in their schooling appeared to have received a lower dosage of *City Connects* since the data were censored once the event of interest had occurred. Inevitably, this made the dosage level and the effect of *City Connects* participation appear inversely related.

Fourth, in this study, it was not possible to empirically establish that censoring was independent of the focal outcome. Censoring can be considered independent if it occurred either because the student transferred out of the district before the end of fifth grade or data was cut-off because BPS data available for this study only encompassed school years 2002 through 2013. For example, for a student who was in third grade in school year 2013, data was not available for

subsequent grades. However, there are other causes of censoring that may be related to the timing of exit from LEP status. Nonetheless, as is common in such studies, it is necessary to assume independence of censoring for the life-table computations. Thus, readers should be aware that the results from the life-table analysis might be biased to some extent.

Fifth, this study found that median lifetimes often exceeded the duration of the study, which was the five years spanning the elementary grades. Thus, it is important to note that the plausible changes in school environment and policies from elementary to middle grades may impact the time to exiting LEP status in ways that this study was not able to measure. As a result, the median time to the event that goes beyond five years may be biased to some extent.

Sixth, as previously mentioned, the sensitivity analysis presented here was exploratory in nature, and a more comprehensive sensitivity analysis could be conducted. To accomplish this, a greater number of simulation trials could be carried out with varying degrees of U s and β_U s. Moreover, the sensitivity analysis could be repeated in parallel with sub-samples 1 and 2, thereby allowing evidence of a convergence in findings with somewhat different *City Connects* samples to be established.

Seventh, this study was not able to examine the precise mechanisms within the intervention that accounted for the treatment effects observed (Dearing et al., 2016). As Dearing et al. state, “Qualitative work examining child, family, school, and community agency experiences with school based student support interventions is critical for understanding how, when, and for whom these interventions are most effective” (Dearing et al., 2016, pg. 894). Even though this study was not able to address this particular limitation, the work presented here nonetheless provided evidence of the value of the *City Connects* intervention.

To date, little research has viewed a student's exiting LEP status as a form of academic success. Although this study aimed to fill this gap, with regard to the effect that *City Connects* participation has on exiting LEP status, studying the degree to which achieving this academic milestone impacts other aspects of LEP students' academic achievements in schools is also important. Examples are report card grades, standardized test scores in elementary through middle schools, retention, dropout, and graduation rates, and high school course-taking patterns (e.g., participation in advanced placement courses and performance on advanced placement exams). More studies can be devised to compare these and other aspects of LEP student participation in *City Connects* with the corresponding aspects of their peers attending comparison schools.

Moreover, since the need for sensitivity analysis stems from the quasi-experimental nature of the research concerning *City Connects* participation, this study can be repeated once data from more recent years that were not included in the study becomes available (i.e., BPS data for school years 2014-15 through 2016-17). Additionally, a similar sub-sample analysis can be conducted with newer *City Connects* cohorts. Likewise, given that *City Connects* has expanded into other school districts in recent years, similar studies can be conducted with data from these districts. Arguably, empirical evidence generated from multiple studies would strengthen the credibility of the causal claims regarding the effectiveness of the *City Connects* intervention.

Finally, future studies should concentrate not only on academic outcomes but also other outcomes, such as students' self-confidence, motivation, and self-control. Such an effort would help enhance understanding of LEP students' psychological needs and strengths, and, thus, could enable *City Connects* to provide them with more targeted community-based prevention, intervention, and enrichment programs.

5.3 Conclusions and Policy Implications

The pattern of results from this study indicate that LEP students receiving the *City Connects* intervention are significantly more likely to exit LEP status earlier than their peers in comparison schools. These findings have practical importance, as the study's models found that approximately 10 percent more *City Connects* students exited LEP status by the end of fifth grade than did non-*City Connects* students. Thus, *City Connects* students were found to be more likely to meet the LEP reclassification criteria by demonstrating academic success and readiness to thrive in mainstream classrooms than their non-*City Connects* counterparts. From these findings can be drawn several implications for policy and practice concerning closing achievement gaps and improving educational opportunities for students living in poverty.

In the U.S., policymakers, educators, and researchers are looking for solutions to improve educational opportunities for students living in poverty with the aim of improving their future economic and social well-being. Within this context, recent years have seen changes and reforms in educational policies and practices designed to better prepare students for college and career by the time they graduate from high school. In the research base of college and career readiness, the most powerful predictor of college and career readiness has been found to be the completion of the high school core curriculum (Achieve Inc., 2004; Conley, 2007; Hein et al., 2013). However, prior research has also shown LEP students to be less likely to complete the core curricula in mathematics and science in high school compared to their English-proficient peers (Aud et al., 2012). One underlying reason is that these students are not considered ready to participate in core academic courses if they have not yet exited LEP status.

This study, however, demonstrated that LEP students who attended high-poverty and urban elementary schools exited LEP status at a younger age when they were provided with

targeted services that addressed out-of-school barriers to learning. In other words, *City Connects* cleared the path for academic success for these students. Thus, policymakers and practitioners should consider the positive implications of a well-designed integrated student support system such as the one *City Connect* provides and implement educational policies that allow such support systems to become common practice.

Policymakers, researchers, and educators are also concerned with the benefit-cost ratios of interventions that will close achievement gaps and improve educational opportunities.

Bowden et al. (2015) conducted a benefit-cost analysis for all the *City Connects* students and discovered the following:

“...[T]he benefit-cost ratio is 3.0 and the net benefits are \$9,280 per student. This result implies that providing the program to a cohort of 100 students over six years would cost society \$457,000 but yield \$1,385,000 in social benefits, for a net benefit of \$928,000. Even under the most conservative assumptions regarding costs and benefits, the program’s benefits exceed its costs. Sensitivity tests show that the benefit-cost ratio lies somewhere between 1 and 11.8, with a best estimate of \$3.00 in benefits per dollar of cost.”

Coupled with the findings of prior research demonstrating the effectiveness of the *City Connects* intervention, i.e., lower retention rates in Grade 6 (Lee-St. John, 2012), increased academic achievement in report card scores and statewide assessments (Walsh et al., 2014), and improved academic achievement for immigrant students (Dearing et al., 2016), the evidence produced by this study shows the *City Connects* intervention to be cost-effective and of practical significance. Consequently, while more research should be done to better understand how, when, and under which conditions such a student support system performs best, policymakers and practitioners should give higher priority to develop, empower, and scale up similar student support systems within the schools.

Finally, income inequality and its impact on the well-being of individuals and the society as a whole have been in the center of recent political debates. Income inequality has been steadily on the rise for the past three decades (OECD, 2009, 2017), and many are concerned that it may also imply greater inequality in the distribution of resources and opportunities (Kirsch & Braun, 2016). One reason that this study focused on LEP students within the context of the *City Connects* intervention was to examine whether a student-support system designed to work in high-poverty urban elementary schools would prove effective for a student subpopulation that was characterized as most at-risk to fail academically since its members faced multiple challenges (Kominski et al., 2011; Sheng et al., 2011). These challenges included acquiring a new language (Artiles & Ortiz, 2002; Kushner & Ortiz, 2000; Zehler et al., 2003), attending inferior schools with low graduation rates (Fry, 2008; Orfield & Lee, 2006), and coming from families with low incomes and lower levels of formal education (Aud et al., 2011, 2012). This study showed that *City Connects* was able to improve the odds of academic success for students who were extremely vulnerable to the risks associated with living in poverty. As such, interventions such as *City Connects* may help lessen the effects of inequality by leveraging community resources to aid the most vulnerable students, thereby increasing their chances of achieving long-term success and well-being.

References

- Abedi, J. (2006). Language issues in item development. In S. M. Downing & T. M (Eds.), *English language acquisition and language instruction educational programs*.
- Abedi, J. (2008a). Classification system for English language learners: Issues and recommendations. *Educational Measurement: Issues and Practice*, 27(3), 17–31.
- Abedi, J. (2008b). Measuring students' level of english proficiency: Educational significance and assessment requirements. *Educational Assessment*, 13, 193–214.
- Achieve Inc. (2004). *Ready or not: Creating a high school diploma that counts--The American diploma project*. Washington, DC: Achieve, Inc.
- Alliance for Excellent Education. (2011). *Saving now and saving later: How high school reform can reduce the nation's wasted remediation dollars*. Washington, D.C.
- Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. In S. Leinhardt (Ed.), *Sociological Methodology* (pp. 61–98). San Francisco: Jossey-Bass.
- An, C. (2015). *Estimating the effectiveness of city connects on middle school outcomes*. (Unpublished doctoral dissertation). Boston College.
- Artiles, A. J., & Ortiz, A. . (2002). *English language learners with special needs: Identification, placement, and instruction*. Washington, DC: Center for Applied Linguistics.
- Aud, S., Hussar, W., Johnson, F., Kena, G., Roth, E., Manning, E., ... Zhang, J. (2012). *The Condition of Education 2012 (NCES 2012-045)*. Washington, DC: National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubsearch>.

- Aud, S., Hussar, W., Kena, G., Bianco, K., Frohlich, L., Kemp, J., & Tahan, K. (2011). *The Condition of Education 2011 (NCES 2011-033)*. Washington, DC: U.S. Government Printing Office.
- Aud, S., Wilkinson-Flicker, S., Kristapovich, P., Rathbun, A., Wang, X., & Zhang, J. (2013). *The Condition of Education 2013 (NCES 2013-037)*. Washington, DC. Retrieved from <http://nces.ed.gov/pubsearch>.
- August, D., & Hakuta, K. (1998). *Educating language-minority children*. Washington, DC: National Academy Press. Retrieved from <http://files.eric.ed.gov/fulltext/ED420737.pdf>
- August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Barton, P. E., & Coley, R. J. (2009). *Parsing the achievement gap II*. Retrieved from <https://www.ets.org/Media/Research/pdf/PICPARSINGII.pdf>
- Bauman, J., Boals, T., Cranley, E., Gottlieb, M., & Kenyon, D. (2007). Assessing comprehension and communication in English state to state for english language learners (ACCESS for ELLs®). In J. Abedi (Ed.), *English language proficiency assessment in the nation: Current status and future practice* (pp. 81–92). Davis: The Regents of the University of California. Retrieved from http://s3.amazonaws.com/academia.edu.documents/18648592/elp_report_ucdavis.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1485196524&Signature=rUHHwi3UH30wfO0%2FmIBXeqMp%2F8E%3D&response-content-disposition=inline%3Bfilename%3DAssessing_comprehension_an

- Berliner, D. C. (2013). Effects of inequality and poverty vs. teachers and schooling on America's youth. *Teachers College Record*, 115(December), 1–26. Retrieved from <http://www.tcrecord.org/Content.asp?ContentId=16889>
- BPS. (2015). *2015–2016 Guide to the Boston Public Schools for Families and Students*.
- Bronfenbrenner, U. (2009). *The ecology of human development*. Harvard University Press.
- Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *The Future of Children*, 7(2), 55–71. Retrieved from <http://www.jstor.org/stable/1602387>
- Bryk, A. S., & Weisberg, H. . (1977). Use of the nonequivalent control group design when subjects are growing. *Psychological Bulletin*, 84, 950–62.
- California Law. Proposition 227: English language in public schools (1998). Retrieved from <http://primary98.sos.ca.gov/VoterGuide/Propositions/227text.htm>
- Carnevale, A. P., Smith, N., & Strohl, J. (2010). *Help Wanted: Projections of jobs and education requirements through 2018*.
- Chester, M. D. (2013). *Transitional guidance on identification, assessment, placement, and reclassification of English language learners*.
- Chomsky, N. (1968). Language and the mind. *Psychology Today*, 1(9), 48–68.
- City Connects. (2014). *The impact of City Connects: Progress report 2014*. Chestnut Hill, MA: Center for Optimized Student Support, Lynch School of Education, Boston College.
- Conley, D. T. (2007). *Redefining college readiness, Volume 3*. Eugene, OR.

- Conteh-Morgan, M. (2002). Connecting the dots: Limited English proficiency, second language learning theories, and information literacy instruction. *The Journal of Academic Librarianship*, 28(4), 191–196.
- Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependency, the optimum age question, and some matters. *Working Papers on Bilingualism*, 19, 121–29.
- Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics*, 11(2).
- Davies, H. T. O., Crombie, I. K., & Tavakoli, M. (1998). When can odds ratios mislead? *British Medical Journal*, 28(316(7136)), 989–91.
- Dearing, E. (2008). Psychological costs of growing up poor. *Annals of the New York Academy of Sciences*, 1136, 324–332. <http://doi.org/10.1196/annals.1425.006>
- Dearing, E., Walsh, M. E., Sibley, E., Lee-St John, T., Foley, C., & Raczek, A. E. (2016). Can community and school-based supports improve the achievement of first-generation immigrant children attending high-poverty schools? *Child Development*, 0(0), 1–15. <http://doi.org/10.1111/cdev.12507>
- DESE. (2013a). Transitional guidance on identification, assessment, placement, and reclassification of english language learners.
- DESE. (2013b). WIDA English language development standards implementation guide (Part I).
- Devaney, B. L., Ellwood, M. R., & Love, J. M. (1997). Programs that mitigate the effects of poverty on children. *The Future of Children*, 7(2), 88–112.

- Diaconu, D. V. (2012). *Modeling science achievement differences between single-sex and coeducational schools: Analyses from Hong Kong, SAR and New Zealand from TIMSS 1995, 1999, and 2003. (Unpublished doctoral dissertation)*. Boston College. Retrieved from ProQuest
- EDFacts. (2016). EDFacts data groups 695 and 696, School year 2014–15. Retrieved from https://nces.ed.gov/ccd/tables/acgr_re_and_characteristics_2014-15.asp
- Educational Opportunities Act (1974).
- Evans, G. W. (2004). The environment of childhood poverty. *The American Psychologist*, 59(2), 77–92. <http://doi.org/10.1037/0003-066X.59.2.77>
- Fast, M., Ferrara, S., & Conrad, D. (2004). *Current efforts in developing English language proficiency measures as required by NCLB: Description of an 18-state collaboration*. Washington, DC.
- Flores, E., Painter, G., Harlow-Nash, Z., & Pachon, H. (2009). *Quepasa?: Are English language-learning students remaining in English learning classes too long?* Policy Brief.
- Francis, D. J., & Rivera, M. O. (2007). Principles underlying English language proficiency tests and academic accountability for ELLs. In J. Abedi (Ed.), *English Language Proficiency Assessment in the Nation: Current Status and Future Practice*. The Regents of the University of California. Retrieved from http://education.ucdavis.edu/sites/main/files/ELP_Report.pdf
- Fry, R. (2008). *The role of schools in the English language learner achievement gap*. Washington, DC: Pew Hispanic Center.

- Gandara, P. (2000). In the aftermath of the storm: English learners in the post-227 era. *Bilingual Research Journal*, 24(1&2), 1–13.
- Garcia, O. (1999). Educating Latino high school students with little formal schooling. In C. Faltis & P. Wolfe (Eds.), *So much to say. Adolescents, bilingualism and ESL in the secondary school* (pp. 61–82). New York: Teachers College Press.
- Garcia, S. B., & Ortiz, A. A. (1988). *Preventing inappropriate referrals of language minority students to special education. Occasional papers in bilingual education*. NCBE New Focus.
- Genesee, F. (2006). *Educating English language learners: A Synthesis of research evidence*.
- Gersten, R., & Baker, S. (2000). What we know about effective instructional practices for English language learners. *Exceptional Children*, 66(4), 454–70.
- Ginsburg, A., & Smith, M. (2016). *Do randomized controlled trials meet the “gold standard”?* American Enterprise Institute.
- Grissom, J. B. (2004). Reclassification of English learners. *Education Policy Analysis Archives*, 12(36), 1–36.
- Guo, S. (2009). *Survival analysis*. Oxford Scholarship Online. Retrieved from <http://www.oxfordscholarship.com.proxy.bc.edu/view/10.1093/acprof:oso/9780195337518.001.0001/acprof-9780195337518>
- Guo, S., & Fraser, M. W. (2010). *Propensity score analysis: Statistical methods and applications*. Thousand Oaks, CA: SAGE.
- Hakuta, K., Butler, Y. G., & Witt, D. (2000). *How long does it take English learners to attain*

- proiciency? Policy Report*. Stanford, CA.
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in highstakes testing. *Educational Measurement: Issues and Practice*, 23(1), 17–27.
- Harder, V. S., Stuart, E. a, & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods*, 15(3), 234–49. <http://doi.org/10.1037/a0019623>
- Hein, V., Smerdon, B., & Sambolt, M. (2013). *College & career readiness & success*.
- Jong, E. J. de. (2004). After exit: Academic achievement patterns of former English language learners. *Educational Policy Analysis Archives*, 12(50).
- Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of American Statistical Association*, 53(282). Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/01621459.1958.10501452>
- Keaton, P. (2012). *Numbers and Types of Public Elementary and Secondary Schools From the Common Core of Data: School Year 2010–11 (NCES 2012-325rev)*. Washington, DC: National Center for Education Statistics. Retrieved from <https://nces.ed.gov/pubs2012/2012325rev.pdf>
- Kieffer, M. J., Lesaux, N. K., & Snow, C. E. (2008). Promises and pitfalls: Implications of NCLB for identifying, assessing, and educating English language learners. In G. L. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, & school reform*. California: Thousand Oaks. Kindler, 2002.
- Kieffer, M. J., & Parker, C. (2017). *Graduation outcomes of students who entered New York City*

- public scholls in grade 5 or 6 as English learner students*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. Retrieved from https://ies.ed.gov/ncee/edlabs/regions/northeast/pdf/REL_2017237.pdf
- Kindler, A. L. (2002). *Survey of the states' limited English proficient students and available educational programs and services: 2000–2001 summary report*. Washington, DC. Retrieved from <http://www.ncele.gwu.edu/policy/states/reports/seareports/0001/sea0001.pdf>
- Kirsch, I., & Braun, H. (Eds.). (2016). Opportunity in America: Setting the stage. In *The dynamics of opportunity in america: Evidence and perspectives*. New York: Springer.
- Kominski, R., Jamieson, A., & Martinez, G. (2011). *At risk conditions of U.S. school age children*.
- Korn, E. L., & Graubard., B. I. (1990). Simultaneous testing of regression coefficients with complex survey data: Use of Bonferroni t statistics. *American Statistician*, 44, 270–276. Retrieved from <http://www.stata.com/support/faqs/statistics/chi-squared-test/>
- Krashen, S. D. (1988). *Second language acquisition and second language learning*. New York: Prentice-Hall International.
- Kushner, M. I., & Ortiz, A. A. (2000). The preparation of early childhood education teachers for English language learners. In *New teachers for a new century: the future of early childhood professional development*. Washington, DC: U.S. Department of Education, National Institute on Early Childhood Development and Education.
- Lau v. Nichols. (1974). *U.S.*

- Lean, D. S., & Colucci, V. A. (2010). *Barriers to learning: The case for integrated mental health services in schools*. R&L Educatio.
- Lee-St. John, T. (2012). *The causal impact of systematic, elementary school student support on middle school grade retention (Unpublised doctoral dissertation)*. Boston College.
- Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin*, 126(2), 309–337. <http://doi.org/10.1037/0033-2909.126.2.309>
- Lieten, G. K. (2008). *Children, structure and agency: Realities across the developing world*. New York, NY: Routledge. Retrieved from <https://books.google.com/books?id=3OGPAgAAQBAJ&pg=PA7&lpg=PA7&dq=%22children+still+suffer+from+poverty%22&source=bl&ots=Y9WDNlf3z2&sig=nPezA97pzOyohhZPJVuFZuldSGU&hl=en&sa=X&ei=cIkRVfKrNq7LsAT184DQAQ&ved=0CC8Q6AEwAw#v=onepage&q=%22children still suffer fr>
- Linquanti, R. (2001). *The redesignation dilemma: Challenges and choices in fostering meaningful accountability for english learners*.
- Logan, K. R., Hansen, C. D., Nieminen, P. K., Wright, E. H., Conference, S., September, I., ... Heath, E. (2015). Student support teams : Helping students succeed in general education classrooms or working to place students in special education? *Education and Training in Mental Retardation and Developmental Disabilities*, 36(3), 280–292. Retrieved from http://www.jstor.org/stable/23879981?seq=1#page_scan_tab_contents
- Mahoney, K. S., & MacSwan, J. (2005). Reexamining identification and reclassification of

- English language learners: A critical discussion of select state practices. *Bilingual Research Journal*, 29(1), 31–42.
- Mavrogordato, M. E. C. (2012). *An An event history analysis examining the rate of reclassification for english language learners (Doctoral dissertation)*. Vanderbilt University.
- McCabe, R. (2000). No one to waste: A report to public decision-makers and community college leaders. In C. Shults (Ed.), *Institutional policies and practices in remedial education: A national study of community colleges*. Washington, DC: American Association of Community Colleges.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23.
- Mill, J. (1843). *A system of logic*.
- Montgomery, M. R., Richards, T., & Braun, H. I. (1986). Child health, breast-feeding, and survival in Malaysia: A random-effects logit approach. *Journal of the American Statistical Association*, 81(394), 297–309.
- NAEP. (2012a). *2004, 2008 and 2012 Long-Term Trend Mathematics Assessments*. Retrieved from http://nationsreportcard.gov/ltt_2012/age17m.aspx#0-6
- NAEP. (2012b). *2004, 2008 and 2012 Long-Term Trend Reading Assessments*. Retrieved from http://nationsreportcard.gov/ltt_2012/age17r.aspx#0-6
- National Conference of State Legislatures. (2015). *Summary of the every student succeeds act, legislation reauthorizing the elementary and secondary education act*.

- NCELA. (2006). FAQ: What legal obligations do schools have to English language learners (ELLs)? Retrieved from <http://www.ncela.us/faqs/view/6>
- NCELA. (2007). FAQ: Which tests are commonly used to determine English language proficiency?
- New York City Board of Education, B. D. of, & Accountability and Assessment. (2000). *ELL subcommittee research studies progress report*.
- No Child Left Behind Act (2001). Retrieved from <http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf>
- OECD. (2009). *Inequality and well-being in OECD countries: What do we know?*
- OECD. (2017). *OECD economic outlook: Better but not good enough*.
- Oller, J. (1977). *Language tests at school: A pragmatic approach*. London: Longman.
- Orfield, G., & Lee, C. (2006). *Racial transformation and the changing nature of segregation*. Cambridge, MA: The Civil Rights Project at Harvard University.
- Parrish, T. B., Merickel, A., Pérez, M., Linquanti, R., Socias, M., Spain, A., ... Delancey, D. (2006). *Effects of the implementation of Proposition 227 on the education of English learners, K-12: Findings from a five-year evaluation*. Washington, D.C.: American Institutes for Research.
- Ramsey, C. A., & Wright, E. N. (1974). Age and second language learning. *The Journal of Social Psychology, 94*, 115–121.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and data*

analysis methods (2nd ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & Toit, M. du. (2011). *HLM 7:*

Linear and nonlinear modeling. Scientific Software International. Retrieved from

http://www.ssicentral.com/hlm/help6/faq/Comparing_non-linear_models_in_HLM.pdf

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor:

New evidence and possible explanations. In *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 91–116). Russell Sage Foundation.

Reutzel, D. R., & Cooter, R. B. (2007). *Strategies for reading assessment and instruction helping*

every child succeed. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.

Rogers, W. H. (1993). Regression standard errors in clustered samples. *Stata Technical Bulletin*,

13(19–23).

Rosenbaum, P. R. (1989). Optimal matching for observational studies. *Journal of the American*

Statistical Association, *84*(408), 1024–1032.

Rosenbaum, P. R., & Rubin, D. B. (1983a). Assessing sensitivity to an unobserved binary

covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society. Series B (Methodological)*, 212–218.

Rosenbaum, P. R., & Rubin, D. B. (1983b). The central role of the propensity score in

observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Rossell, C. H. (2000). *Different questions, different answers: A critique of the Hakuta, Butler,*

and Witt report, "How long does it take English learners to attain proficiency?"

- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, *125*(1), 175–214.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.
- Rubin, D. B. (1986). Statistics and causal inference: comment: Which ifs have causal answers. *Journal of the American Statistical Association*, *81*(396), 961–962.
- Rubin, D. B. (1990). Formal modes of statistical inference for causal effects. *Journal of Statistical and Inference*, *25*, 279–292.
- Sánchez, M. T., Ehrlich, S., Midouhas, E., & O'Dwyer, L. (2009). *Analyzing performance by grade 10 Hispanic high school students on the Massachusetts state assessment (Issues & Answers Report, REL 2009–No. 071)*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands. Retrieved from <http://files.eric.ed.gov/fulltext/ED505420.pdf>
- Sánchez, M. T., Parker, C., Akbayin, B., & McTigue, A. (2010). *Processes and challenges in identifying learning disabilities among students who are English language learners in three New York State districts*. Washington, DC: U.S. Retrieved from <http://ies.ed.gov/ncee/edlabs>
- Sastry, N., & Pebley, A. R. (2010). Family and neighborhood sources of socioeconomic inequality in children's achievement. *Demography*, *47*(3), 777–800.
<http://doi.org/10.1353/dem.0.0114>

- Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs (report from the Governing Board of the American Educational Research Association Grants Program)*. Washington, DC:
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth, Cengage Learning.
- Sheng, Z., Sheng, Y., & Anderson, C. J. (2011). Dropping out of school among ELL students: Implications to schools and teacher education. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 84(3), 98–103.
- Singer, J. D., & Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics*, 18, 155–195.
- Slama, R. B. (2012). *A roadmap for true accountability: Reconceptualizing language-learning services, reclassification practices and monitoring systems for english language learners in U.S. public schools (Doctoral dissertation)*. Harvard University.
- Smeeding, T. M. (2016). Gates, gaps, and intergenerational mobility: the importance of an even start. In I. Kirsch & H. Braun (Eds.), *The dynamics of opportunity in america: Evidence and perspectives* (1st ed., pp. 255–295). Springer.
- Snijder, T., & Bosker, R. (1999). *Multilevel analysis*. London: SAGE.
- Sparks, D., & Malkus, N. (2013). *First-Year Undergraduate Remedial Coursetaking: 1999–2000, 2003–04, 2007–08 (NCES 2013–013)*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved

from <http://nces.ed.gov/pubs2013/2013013.pdf>

Stefanakis, E. H. (1998). *Whose judgment counts? Assessing bilingual children, K-3.*

Portsmouth, NH: Heinemann.

Texas Education Agency. (2016). Language proficiency assessment committee. Retrieved

February 11, 2016, from <http://programs.esc20.net/default.aspx?name=lpac.home>

The Conference Board. (2006). *Are they really ready to work?: Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st century U.S. workforce.*

Thompson, K. D. (2012). *Are we there yet? Exploring English learners' journey to reclassification and beyond.* Stanford University, Palo Alto, CA.

Tjur, T. (2009). Coefficients of determination in logistic regression models—A new proposal:

The coefficient of discrimination. *The American Statistician*, 63(4).

U.S. Census Bureau. (2012). Most children younger than age 1 are minorities, census bureau reports. Retrieved from

<http://www.census.gov/newsroom/releases/archives/population/cb12-90.html>

U.S. Census Bureau. (2013). American community survey. Retrieved from

http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_13_1YR_S1701&prodType=table

U.S. Census Bureau. (2015). State and county quick facts. Retrieved from

<http://quickfacts.census.gov/qfd/states/00000.html>

U.S. Department of Education. (2005). Developing programs for english language learners:

- HEW memorandum. Retrieved January 1, 2005, from <http://www2.ed.gov/about/offices/list/ocr/ell/may25.html>
- U.S. Department of Education. (2013). *What works clearinghouse: Procedures and standards handbook (version 3.0)*. Retrieved from <http://whatworks.ed.gov>
- U.S. Department of Education. (2014a). *The Condition of Education 2014 (NCES 2014-083)*. Retrieved from <https://nces.ed.gov/fastfacts/display.asp?id=96>
- U.S. Department of Education. (2014b). *WWC evidence review protocol for dropout prevention interventions, version 3.0*.
- U.S. Department of Justice. (1964). Types of educational opportunities discrimination. Retrieved January 1, 2015, from <http://www.justice.gov/crt/about/edu/types.php>
- U.S. Department of Justice. (2010). Settlement agreement between the UNITED STATES of america and the Boston public schools. Retrieved from <http://www.justice.gov/crt/about/edu/documents/bostonsettle.pdf>
- Walsh, M. E., & DePaul, J. (2001). The essential role of school-community partnerships in school counseling. In H. L. K. Coleman & C. Yeh (Eds.), *Handbook of School Counseling* (pp. 765–784). Routledge.
- Walsh, M. E., Madaus, G. F., Raczek, A. E., Dearing, E., Foley, C., An, C., ... Beaton, A. (2014). A new model for student support in high-poverty urban elementary schools: Effects on elementary and middle school academic outcomes. *American Educational Research Journal, 51*(4), 704–737. <http://doi.org/10.3102/0002831214541669>
- WIDA. (2015). Consortium members. Retrieved from <https://www.wida.us/membership/states/>

World Health Organization. (2015). Child health. Retrieved January 1, 2015, from http://www.who.int/topics/child_health/en/

Yamaguchi, K. (1991). *Event history analysis*. Newbury Park, Calif.: SAGE Publications.

Yoshikawa, H., Aber, J. L., & Beardslee, W. R. (2012). The effects of poverty on the mental, emotional, and behavioral health of children and youth: Implications for prevention. *American Psychologist*, 67(4), 272–284. <http://doi.org/10.1037/a0028015>

Zehler, A. M., Fleischman, H. L., Hopstock, P. J., T.G., S., Pendzick, M. L., & Sapru, S. (2003). *Descriptive study of services to LEP students and LEP students with disabilities (No.4 Special topic report: findings on special education LEP students)*. Arlington, VA: Development Associates.

Appendix A

We can simplify Equation 3.13 to get Equation 3.14:

1. Distribute the parenthesis

$$\begin{aligned}
 l &= \sum_{i=1}^n [(1 - c_i) \log h_{ij_i} + (1 - c_i) \sum_{j=1}^{j_i-1} \log(1 - h_{ij})] + c_i \sum_{j=1}^{j_i} \log(1 - h_{ij}) \\
 l &= \sum_{i=1}^n [(1 - c_i) \log h_{ij_i} + \sum_{j=1}^{j_i-1} \log(1 - h_{ij}) - c_i \sum_{j=1}^{j_i-1} \log(1 - h_{ij}) \\
 &\quad + c_i \sum_{j=1}^{j_i} \log(1 - h_{ij})]
 \end{aligned}$$

2. Write the first term for $\sum_{j=1}^{j_i} \log(1 - h_{ij})$ so that the sum is in the form of j_{i-1} to use in cancellation in step

$$l = \sum_{i=1}^n [(1 - c_i) \log h_{ij_i} + \sum_{j=1}^{j_i-1} \log(1 - h_{ij}) - c_i \sum_{j=1}^{j_i-1} \log(1 - h_{ij}) + c_i \log(1 - h_{ij}) + c_i \sum_{j=1}^{j_{i-1}} \log(1 - h_{ij})]$$

3. Same terms with opposite signs cancel each other.

$$\begin{aligned}
 l &= \sum_{i=1}^n [(1 - c_i) \log h_{ij_i} + \sum_{j=1}^{j_i-1} \log(1 - h_{ij}) + c_i \log(1 - h_{ij}) - \sum_{j=1}^{j_{i-1}} \log(1 - h_{ij}) \\
 &\quad + \sum_{j=1}^{j_{i-1}} \log(1 - h_{ij})] \\
 l &= \sum_{i=1}^n [(1 - c_i) \log h_{ij_i} + \sum_{j=1}^{j_i-1} \log(1 - h_{ij}) + c_i \log(1 - h_{ij})]
 \end{aligned}$$

4. Add and subtracting the same term of $\sum_{j=1}^{j_i} \log(1 - h_{ij})$

$$l = \sum_{i=1}^n [(1 - c_i) \log h_{ij_i} + \sum_{j=1}^{j_i-1} \log(1 - h_{ij}) + c_i \log(1 - h_{ij}) - \sum_{j=1}^{j_i} \log(1 - h_{ij}) + \sum_{j=1}^{j_i} \log(1 - h_{ij})]$$

5. Write the first term such that sum is in the form of j_{i-1} for cancellation

$$l = \sum_{i=1}^n [(1 - c_i) \log h_{ij_i} + \sum_{j=1}^{j_{i-1}} \log(1 - h_{ij}) + c_i \log(1 - h_{ij}) - \log(1 - h_{ij}) - \sum_{j=1}^{j_{i-1}} \log(1 - h_{ij}) + \sum_{j=1}^{j_i} \log(1 - h_{ij})]$$

6. Take into $\log(1 - h_{ij})$ parenthesis

$$l = \sum_{i=1}^n [(1 - c_i) \log h_{ij_i} + c_i \log(1 - h_{ij}) - \log(1 - h_{ij}) + \sum_{j=1}^{j_i} \log(1 - h_{ij})]$$

7. Rearrange the highlighted part

$$l = \sum_{i=1}^n [(1 - c_i) \log h_{ij_i} - (1 - c_i) \log(1 - h_{ij}) + \sum_{j=1}^{j_i} \log(1 - h_{ij})]$$

$$l = \sum_{i=1}^n [(1 - c_i) \log \left(\frac{h_{ij_i}}{1 - h_{ij}} \right) + \sum_{j=1}^{j_i} \log(1 - h_{ij})] \quad (3.1)$$

Appendix B

We can also define the outcome of y_{ij} based on whether the individual is censored or not.

If the individual is not censored, then $c_i = 0$ and y_{ij} is equal to one only for the last period and zero for all the earlier periods (Singer & Willett, 1993). Similarly, if the individual is censored, then $c_i = 1$ and y_{ij} is equal to zero for all the time periods, including the very last one (Singer & Willett, 1993). Thus, we can express this as:

$$\sum_{j=1}^{j_i} y_{ij} = (1 - c_i) = \begin{cases} 1 & \text{when } c_i = 0 \\ 0 & \text{when } c_i = 1 \end{cases}$$

Equation 3.15 can be re-written as the following by multiplying both sides with the same term of $\log\left(\frac{h_{ij_i}}{1-h_{ij}}\right)$:

$$\sum_{j=1}^{j_i} y_{ij} \log\left(\frac{h_{ij}}{1-h_{ij}}\right) = (1 - c_i) \log\left(\frac{h_{ij_i}}{1-h_{ij}}\right)$$

We can then substitute right side of Equation 3.16 into Equation 3.14. Then, with rearranging and collecting the like terms, Equation 3.14 becomes:

1. $l = \sum_{i=1}^n [\sum_{j=1}^{j_i} y_{ij} \log\left(\frac{h_{ij}}{1-h_{ij}}\right) + \sum_{j=1}^{j_i} \log(1 - h_{ij})]$
2. $l = \sum_{i=1}^n \sum_{j=1}^{j_i} [y_{ij} \log\left(\frac{h_{ij}}{1-h_{ij}}\right) + \log(1 - h_{ij})]$
3. $l = \sum_{i=1}^n \sum_{j=1}^{j_i} [y_{ij} \log h_{ij} - y_{ij} \log(1 - h_{ij}) + \log(1 - h_{ij})]$

$$4. \quad l = \sum_{i=1}^n \sum_{j=1}^{j_i} [\log h_{ij}^{y_{ij}} - \log(1 - h_{ij})^{y_{ij}} + \log(1 - h_{ij})]$$

$$5. \quad l = \sum_{i=1}^n \sum_{j=1}^{j_i} [\log h_{ij}^{y_{ij}} + \left(\frac{\log(1-h_{ij})}{\log(1-h_{ij})^{y_{ij}}}\right)]$$

$$l = \sum_{i=1}^n \sum_{j=1}^{j_i} [\log h_{ij}^{y_{ij}} + \log(1 - h_{ij})^{(1-y_{ij})}]$$

Then, by antilogging Equation 3.17, the likelihood function becomes:

$$L = \prod_{i=1}^n \prod_{j=1}^{j_i} h_{ij}^{y_{ij}} (1 - h_{ij})^{(1-y_{ij})}$$