

**TIME TO PROFICIENCY IN YOUNG ENGLISH LEARNERS AND FACTORS THAT  
AFFECT THE TIME**

By

Xiaowan Zhang

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Second Language Studies—Doctor of Philosophy

2021

## **ABSTRACT**

### **TIME TO PROFICIENCY IN YOUNG ENGLISH LEARNERS AND FACTORS THAT AFFECT THE TIME**

By

Xiaowan Zhang

English learner (EL) children (i.e., children learning English as a second language) constitute one of the fastest growing, yet disproportionately underachieving, segments of the U.S. public-school population. A main difficulty that ELs face at school is learning English and content-area knowledge in tandem. Low levels of English proficiency may limit ELs' abilities to benefit from content instruction in English or to demonstrate knowledge and skills on mainstream academic assessments (Cook, Linn, & Jung, 2012). This study investigates the time it takes for EL children to attain English proficiency and factors that affect the time by drawing on longitudinal EL data from Michigan. Socio-demographic and English-proficiency assessment data were requested from the Michigan Education Research Institute (MERI) on six cohorts of students who entered Michigan public schools in kindergarten as ELs in 2013-2014 through 2018-2019. All students were followed for up to six years from kindergarten through fifth grade. Discrete-time survival analysis was used to estimate the time that ELs took to attain proficiency as measured by Michigan's state English language proficiency assessment (i.e., ACCESS for ELLs) and to explore the relationship between time to proficiency and six factors of interests: (a) primary disability type, (b) primary home language, (c) poverty status, (d) home English use, (e) instructional programming, and (f) retention. Findings showed that half of ELs who entered Michigan public schools in kindergarten attained proficiency in five years, with writing being the largest barrier to proficiency for those students. Findings of this study further indicated ELs' time to proficiency was significantly related to their

sociodemographic and educational backgrounds. Specifically, ELs with disabilities, ELs speaking some particular home languages (e.g., Arabic, AfroAisatic, and Spanish), and ELs who ever lived in poverty were less likely to attain proficiency than their peers who did not have those backgrounds. In addition, ELs who received partial instruction in their home language (L1) and ELs who never used English at home were equally likely, and in some cases slightly more likely, to attain proficiency as compared with their peers who were immersed full-time in an English-only environment at school and at home. Lastly, students who were ever retained in grade were less likely to attain proficiency than their peers who were never retained in upper elementary grades despite short-term gains associated with retention in early elementary grades. Implications of these findings point to the importance of improving current accountability systems to reflect the diversity of the EL population, supporting L1 maintenance in the home and school contexts, monitoring the effect of retention on ELs, and providing better writing instruction for ELs.

Copyright by  
XIAOWAN ZHANG  
2021

## ACKNOWLEDGEMENTS

It is hard to believe that my PhD journey at Michigan State University has almost come to an end. Before I say goodbye to this beautiful campus that I call home for five years, I would like to thank everyone who has made this journey so joyful and rewarding for me.

First, I want to thank my advisor and the chair of my dissertation committee, Dr. Paula Winke, for her help and support throughout the past five years. I still vividly remember the first day I met Paula. In her office in Wells Hall, Paula welcomed me with a big, warm smile. That smile melted my heart and swept away all the fears I had about the new life in Michigan and the sickness from being far away from home. After learning that it was my first day on campus, Paula put down her work and took me for a campus tour. It was a warm summer day, and every other day I spent with Paula has been as warm and summery as the first day we met. Paula is an amazing researcher, teacher, and mentor. I learned so much from her about language testing and policy, conducting research, writing about research, and applying for grants. She was also such a great role model for me, always reminding me that I need to use my research to advocate for the most vulnerable and to promote fairness, equity, and diversity in the society. Paula was the reason why I started this PhD journey at Michigan State University, and she is also the reason why I wish that this journey has not come to an end so soon.

I would also like to express my sincere gratitude to the other members on my dissertation committee: Dr. Shawn Loewen, Dr. Katharine Strunk, and Dr. Koen Van Gorp. They all provided me with valuable advice on planning and designing this project. I could not have asked for a better team of topical expertise that aligns to this dissertation. I am grateful to Koen and Shawn for helping me strengthen the theoretical foundation of this project from a perspective of

second language acquisition. I am also grateful to Katharine (who is so kind to serve on my committee even though she is not a professor in Second Language Studies) for her generous feedback on the policy and methodological aspects of the project.

In addition, I wish to acknowledge the support from the Language Learning grant, which sponsored my consultation with Dr. Chris Andrews (expert on survival analysis and much more) in Consulting for Statistics, Computing, and Analytics Research at the University of Michigan. I am very appreciative of Chris' advice on every step I took in completing the statistical analysis. I have learned so much from him about data cleaning, variable coding, data analysis, and visualization.

This past year has been difficult for us all. I thank my fellow students (and alumni) for keeping me in a warm SLS community during the pandemic. Especially, I appreciate the moral support from Dylan Burton, Yingzhao Chen, Stella He, Jongbong Lee, Shinye Lee, Matt Kessler, Melody Ma, Myeongeun Son, and Monique Yoder. I also thank Dan Isbell for always being available for my questions.

Last but not least, I want to thank my family. My mom and dad have always been supportive for my choices and decisions. I am extremely lucky to be their daughter. I am also thankful to my dear husband, Zhonghao Wang: Thank you for coming to the United States with me, thank you for taking such good care of me, and thank you for being my listener, friend, and life partner. I could not have done this without you.

## TABLE OF CONTENTS

LIST OF TABLES .....	x
LIST OF FIGURES .....	xiv
CHAPTER 1 INTRODUCTION.....	1
CHAPTER 2 CHARACTERISTICS OF ENGLISH LEARNERS (ELs) .....	5
CHAPTER 3 LEGAL UNDERPINNINGS .....	8
3.1 Current Education Policies Governing the Education of English Learners (ELs) .....	8
3.2 EL Identification, Classification, and Reclassification .....	11
3.2.1 EL Initial Identification .....	12
3.2.2 EL Classification .....	13
3.2.3 EL Reclassification.....	14
CHAPTER 4 TIME TO ENGLISH PROFICIENCY .....	17
CHAPTER 5 FACTORS RELATED TO TIME TO ENGLISH PROFICIENCY .....	32
5.1 Overview of This Study's Six Factors of Interest .....	33
5.2 A Closer Look at This Study's Six Factors of Interest.....	34
5.2.1 Primary Disability Type .....	34
5.2.2 Primary Home Language.....	36
5.2.3 Home English Use .....	37
5.2.4 Poverty Status .....	39
5.2.5 Instructional Programming .....	41
5.2.6 Retention.....	45
CHAPTER 6 CONTEXT FOR THE STUDY .....	49
6.1 EL Identification, Classification, and Reclassification Practices in Michigan .....	50
6.2 EL Instructional Programs in Michigan .....	52
CHAPTER 7 RESEARCH QUESTIONS AND METHOD .....	57
7.1 Research Questions .....	57
7.2 Method.....	58
7.2.1 Data.....	58
7.2.1.1 Instructional Program Participation Patterns.....	64
7.2.1.2 Poverty Patterns.....	68
7.2.1.3 Home-English-Use Patterns .....	69
7.2.1.4 Retention.....	70
7.2.1.5 Information on School Districts .....	71
7.2.2 Measures .....	72
7.2.3 Variables.....	77
7.2.3.1 Outcome Variables .....	77
7.2.3.2 Predictors .....	77

7.2.3.3 Control Covariates .....	80
7.2.4 Analysis .....	83
7.2.4.1 Research Question 1 .....	85
7.2.4.2 Research Question 2 .....	91
7.2.4.3 Research Question 3 .....	94
7.2.4.4 Assumptions Underlying Discrete-time Survival Analysis.....	94
7.2.4.5 Sensitivity Analysis .....	95
CHAPTER 8 RESULTS.....	97
8.1 Research Question 1 .....	97
8.2 Research Question 2 .....	101
8.2.1 Primary Disability Type .....	102
8.2.2 Primary Home Language.....	104
8.2.3 Poverty Status .....	109
8.2.4 Home English Use .....	111
8.2.5 Instructional Program in Kindergarten .....	112
8.2.6 Retention.....	115
8.2.7 Results for the Conditional Discrete-time Hazard Model (Equation 3 in Chapter 7) ..	117
8.3 Research Question 3 .....	129
CHAPTER 9 DISCUSSION .....	137
9.1 Time to Proficiency .....	138
9.2 Factors that Affect Time to Proficiency .....	140
9.2.1 Primary Disability Type .....	140
9.2.2 Primary Home Language.....	142
9.2.3 Poverty Status .....	145
9.2.4 Home English Use .....	146
9.2.5 Instructional Programming .....	148
9.2.6 Retention.....	151
9.3 Barrier to Proficiency .....	153
9.4 Generalizability .....	154
9.5 Is Survival Good or Bad? .....	156
CHAPTER 10 PRACTICAL IMPLICATIONS .....	161
10.1 Improving School Accountability Systems to Reflect the Diversity of the EL Subgroup .....	161
10.2 Taking More Measures to Support ELs' L1 Development .....	163
10.3 Carefully Investigating and Monitoring the Effect of retention.....	166
10.4 Strengthening Writing Instruction for ELs.....	167
CHAPTER 11 LIMITATIONS .....	169
CHAPTER 12 CONCLUSION .....	173
APPENDICES .....	175
APPENDIX A Data Preparation .....	176
APPENDIX B Considerations in Coding Predictors.....	184



APPENDIX C Demographic Characteristics of the Subsample .....	193
APPENDIX D Results for Multicollinearity Checks .....	196
APPENDIX E Assumption Testing.....	198
APPENDIX F. Results for Research Question 1 Based on the Subsample .....	202
APPENDIX G. Model Results for Research Question 2 Based on the Full Sample .....	204
APPENDIX H. Full Model Results for Research Question 2 Based on the Full Sample.....	207
APPENDIX I. Graphing Method Description.....	220
APPENDIX J. Results for Sensitivity Analyses.....	221
APPENDIX K. Results for Research Question 3 Based on the Subsample.....	225
REFERENCES .....	227

## LIST OF TABLES

Table 1. Summary of post-NCLB, longitudinal studies estimating the time that ELs take to become reclassified/attain proficiency .....	19
Table 2. Classification and coding of EL instructional programs provided by Michigan public schools .....	55
Table 3. Mean characteristics of the full analytic sample by cohort .....	60
Table 4. The most common EL program participation trajectories for 2013-2014 kindergarten entrants prior to leaving the study ( $n = 9566$ ) .....	64
Table 5. The most common program participation trajectories for 2013-2014 kindergarten entrants who ever participated in transitional bilingual programs prior to leaving the study ( $n = 1888$ ) .....	65
Table 6. The most common program participation trajectories for 2013-2014 kindergarten entrants who ever participated in dual-language programs prior to leaving the study ( $n = 448$ ) ..	65
Table 7. The numbers and percentages of students in ESL, transitional bilingual, and dual-language bilingual instruction by cohort .....	66
Table 8. The numbers and percentages of ELs who were ever in transitional bilingual and dual-language bilingual programs in 2013-2014 kindergarten entrants and in the entire sample by ten most frequently reported home languages .....	67
Table 9. The ten most common poverty trajectories in 2013-2014 kindergarten entrants prior to leaving the study ( $n = 9566$ ) .....	68
Table 10. The eleven most common home-English-use trajectories in 2013-2014 kindergarten entrants prior to leaving the study ( $n = 9566$ ) .....	69
Table 11. The number of retentions among 2013-2014 kindergarten entrants by grade ( $n = 797$ ) .....	70
Table 12. Descriptive statistics for 2013-2014 kindergarten entrants by retention status ( $n = 9,566$ ) .....	71
Table 13. Mean characteristics for the ten districts with largest EL populations between 2013-2014 and 2018-2019 .....	72
Table 14. Test constructs of the four ACCESS subsections .....	73
Table 15. ACCESS proficiency levels for the full sample by skill, cohort, and academic year ...	75
Table 16. Summary of outcome variables, predictors, and control covariates .....	81

Table 17. Traditional and adjusted survival analysis terminology for social science, general education, and applied linguistics.....	87
Table 18. Life table for proficiency attainment of ELs in the full sample, using data from 2013-2014 through 2018-2019 .....	98
Table 19. Results of the unconditional discrete-time hazard model (specified in Equation 1 in Chapter 7) predicting hazard of the full sample attaining proficiency as a function of time, using data from 2013-2014 through 2018-2019.....	99
Table 20. Life table by primary disability status for the full sample, using data from 2013-2014 through 2018-2019 .....	103
Table 21. Life table by primary home language for the full sample, using data from 2013-2014 through 2018-2019 .....	106
Table 22. Life table by poverty status for the full sample, using data from 2013-2014 through 2018-2019 .....	109
Table 23. Life table by home English use for the full sample, using data from 2013-2014 through 2018-2019 .....	111
Table 24. Life table by instructional program for the full sample, using data from 2013-2014 through 2018-2019 .....	113
Table 25. Life table by retention for the full sample, using data from 2013-2014 through 2018-2019 .....	115
Table 26. Results for the conditional discrete-time hazard model predicting proficiency hazard for students in the subsample, using data from 2013-2014 through 2018-2019 .....	118
Table 27. Life table for attaining individual proficiency criteria for ELs in the full sample, using data from 2013-2014 through 2018-2019.....	130
Table 28. Results of the discrete-time hazard model (specified in Equation 1 in Chapter 7) predicting hazard of the full sample reaching individual proficiency criteria as a function of time, using data from 2013-2014 through 2018-2019 .....	131
Table 29. Traditional and adjusted survival analysis terminology for social science, general education, and applied linguistics.....	158
Table 30. A list of study variables and their origins in the MDE data .....	176
Table 31. A list of study variables and their origins in the CEPI data .....	177
Table 32. A fake case with inconsistent cross-year, within-person information on time-invariant variables.....	178
Table 33. Numbers of cases with inconsistent data on time-invariant variables.....	179

Table 34. The number of students who were censored and the number of students who attained ELP by the number of years since the student entered kindergarten.....	183
Table 35. Disability categories and the numbers of records and students in each category .....	184
Table 36. Results for two program models, Model ProgramTV and Model ProgramTI, coding instructional program as time-varying and time-invariant, respectively.....	187
Table 37. Three methods for coding poverty as a time-varying variable (2 fake examples) .....	188
Table 38. Results for three models that used poverty status and time to predict the outcome of proficiency attainment, each using a different method to code poverty status.....	189
Table 39. Three methods for coding English use at home as a time-varying variable (2 fake examples).....	190
Table 40. Results for three models that used English use at home and time to predict the outcome of proficiency attainment, each using a different method to code English use at home. ....	191
Table 41. Two methods for coding retention as a time-varying variable (2 fake examples) .....	192
Table 42. Results for two models that used retention and time to predict the outcome of proficiency attainment, each using a different method to code retention.....	192
Table 43. Mean demographic characteristics of the subsample by cohort.....	193
Table 44. Polychoric correlation among poverty status, home English use, instructional programming, and retention .....	196
Table 45. Relationship between primary home language and poverty status .....	196
Table 46. The number of students in the full sample who were censored due to mechanisms (a) and (b) by time period .....	198
Table 47. Summary of the results of the discrete-time hazard model predicting attrition .....	199
Table 48. BIC of Models 1 and Model 2 for six predictors of interest .....	201
Table 49. Life table for proficiency attainment of ELs in the sub sample, using data from 2014-2015 through 2018-2019 .....	202
Table 50. Results of the unconditional discrete-time hazard model (specified in Equation 1) predicting hazard of proficiency attainment as a function of time .....	203
Table 51. Results for the conditional discrete-time hazard model predicting proficiency hazard for students in the full sample, using data from 2013-2014 through 2018-2019 .....	205
Table 52. Full results for the conditional discrete-time hazard model predicting proficiency hazard for students in the subsample, using data from 2013-2014 through 2018-2019.....	207

Table 53. Results for three sensitivity analyses in comparison with results of the model used to answer research question 2 (based on subsample data).....222

Table 54. Results of the discrete-time hazard model (specified in Equation 1 in Chapter 7) predicting hazard of the subsample reaching individual proficiency criteria as a function of time, using data from 2013-2014 through 2018-2019 .....225

## LIST OF FIGURES

Figure 1. Changes in the risk set from time 0 (year 1 in school) through time 5 (year 6 in school) .....	89
Figure 2. Model-estimated hazard of proficiency (left panel) and cumulative proportion of proficient students (right panel) by time period, using data of the full sample .....	99
Figure 3. Life-table-derived hazard of proficiency (left panel) and cumulative proportion of proficient students (right panel) by time period and by disability status, using data of the full sample .....	104
Figure 4. Life-table-derived hazard of proficiency (left panel) and cumulative proportion of proficient students (right panel) by time period and by primary home language, using data of the full sample .....	108
Figure 5. Life-table-derived hazard of proficiency (left panel) and cumulative proportion of proficient students (right panel) by time period and by poverty status, using data of the full sample .....	110
Figure 6. Life-table-derived hazard (likelihood) of proficiency (left panel) and cumulative proportion of proficient students (right panel) by time period and by home English use, using data of the full sample .....	112
Figure 7. Life-table-derived hazard of proficiency (left panel) and cumulative proportion of proficient students (right panel) by time period and by instructional programming, using data of the full sample .....	114
Figure 8. Life-table-derived hazard of proficiency (left panel) and cumulative proportion of proficient students (right panel) by time period and by retention, using data of the full sample .....	116
Figure 9. Fitted cumulative proficiency rate by time period and primary disability status, using data of the subsample .....	120
Figure 10. Fitted cumulative proficiency rate by time period and primary home language (including Albanian, Arabic, Bengali, Chinese, Spanish, and Other), using data of the subsample .....	123
Figure 11. Fitted cumulative proficiency rate by time period and primary home language (including AfroAsiatic, Austronesian, BaltoSlavic, Dravidian, IndoIranian, Italic/Germanic, Japanese/Korean), using data of the subsample .....	124
Figure 12. Fitted cumulative proficiency rate by time period and poverty status, using data of the subsample .....	125

Figure 13. Fitted cumulative proficiency rate by time period and home English use, using data of the subsample .....	126
Figure 14. Fitted cumulative proficiency rate by time period and instructional programming, using data of the subsample.....	127
Figure 15. Fitted cumulative proficiency rate by time period and retention, using data of the subsample .....	129
Figure 16. Fitted cumulative proficiency rates for each of the five proficiency criteria (i.e., ACCESS listening, ACCESS speaking, ACCESS reading, ACCESS writing, and ACCESS composite score), using data of the full sample .....	132
Figure 17. Visualization of the data cleaning and preparation process .....	183

## **CHAPTER 1 INTRODUCTION**

There is an imperative need to understand the English language development trajectories of children in the U.S. public schools who speak a language other than English at home and who are in the process of developing their proficiency in English as a second language, referred to here as English learner (EL) students. ELs constitute one of the fastest growing, yet disproportionately underachieving, segments of the U.S. public-school population. A main difficulty that ELs face at school is learning English and content-area knowledge in tandem. While they are in the process of acquiring English, ELs may have a limited ability to benefit from content instruction in English and to demonstrate knowledge and skills on mainstream academic assessments (Cook, Linnquanti, Chinen, & Jung, 2012). For example, researchers have found that Spanish-home-language students' low levels of English, particularly in oral and vocabulary skills, are major obstacles to their acquisition of English reading comprehension (Grimm, Solari, & Gerber, 2018; Kieffer, 2008, 2012; Nakamoto, Lindsey, & Manis, 2007). In addition, evidence suggests that ELs tend to achieve unreliable and potentially invalid low scores on content assessments in English due to the high language demands placed on ELs by those assessments (Abedi, 2004; Bailey & Butler, 2004; Winke & Zhang, 2019).

National and state reports drawing on cross-sectional assessment data have consistently documented achievement gaps between ELs and non-ELs in various content areas, with those gaps typically growing wider at older grade levels (National Center for Education Statistics, 2017a; National Center for Education Statistics, 2017b; Sugarman & Geary, 2018). While short-term lags are admissible for ELs because they are by definition not expected to perform at the same level on mainstream assessments as their English proficient peers (Abedi, 2004; Saunders & Marcelletti, 2013), what is concerning is long-term underachievement of those students who



retain their EL status for prolonged periods of time; these students are the so-called long-term ELs (see Flores & Rosa, 2015, however, for a critique of the term). In educational measurement, long-term ELs usually refer to students who have spent six or more years in U.S. schools without developing sufficient English proficiency needed to perform mainstream academic work in English (Kieffer & Parker, 2016). Researchers have noted that long-term ELs share a number of common characteristics, including low English literacy skills in particular, that place them at elevated risk for low-track placement, grade retention, and drop out in secondary school (Menken & Kleyn, 2010; Menken, Kleyn, & Chae, 2012; Slama, 2012).

Current educational policies hold state and local educational agencies accountable for the academic success of all children, including those in the EL subgroup. Particularly, states must ensure that ELs are making appropriate progress toward English language proficiency (ELP) by providing specialized language-learning services to ELs and establishing targets for ELs' English language development (ELD). Appropriately serving and monitoring ELs toward ELP relies on a good understanding of how long it takes EL children to attain proficiency in English, and of factors that may affect this time. Although there have been a growing number of attempts to understand these issues, key questions remain. For example, the most widely-cited time frame for ELP attainment—4 to 7 years—was provided by Hakuta, Butler, and Witt (2000). As I will explain in the literature review further (see Chapter 4), Hakuta et al. calculated this average by using cross-sectional (rather than longitudinal) data. Despite the study's profound influence in EL education, its time-to-proficiency estimate is limited by the cross-sectional design, where the time effect is confounded with the cohort effect. Researchers using a longitudinal design have primarily focused on EL reclassification, an event that marks a change in a student's status from *EL* to *former EL*. Although reclassification is closely related to ELP attainment, existing

estimates of time to reclassification have limited generalizability to serve as valid proxies for time-to-proficiency because (a) reclassification usually involves considerations irrelevant to the construct of ELP (e.g., content-area assessment, teacher evaluation, monetary incentives) and because (b) these estimates are largely based on outdated, state-dependent definitions and measures of ELP prior to the development of a common definition of ELs by large, multi-state ELP assessment consortia.

In view of these limitations, for this current study on ELs' time to proficiency, I obtain updated, empirically grounded time-to-proficiency estimates using six years of longitudinal EL data from Michigan, a member of the World-Class Instructional Design and Assessment (WIDA; <https://wida.wisc.edu>) Consortium that is currently made up of 40 states, territories, and federal agencies. I estimate ELs' time to proficiency independent of a particular state's reclassification considerations (a novel approach), which will allow for wider extrapolation of the findings. In addition, I examine factors that are associated with faster or slower proficiency attainment in ELs. Compared with previous studies, this study provides estimates of broader generalizability in that the estimates are comparable to those from the other 39 states (or regions) within the WIDA consortium. These estimates also have better validity because this study minimizes considerations irrelevant to the core construct of ELP. As such, the current study has the potential to influence policymakers when they establish and revise ELD targets and design accountability systems for ELs from different backgrounds. It also adds to the limited research on child second language acquisition and factors that may affect children's' rates of acquisition (Paradis, 2008).

This dissertation proceeds in the following structure: In chapter 1 (this current chapter), I provide an overview of the study and its background. Chapter 2 describes the characteristics of

K-12 ELs in the United States. Current educational policies that influence ELs' English language acquisition are briefly reviewed in Chapter 3. Chapters 4 and 5 review previous studies that examined the time that ELs take to reach proficiency and factors that are associated with variation in ELs' time to proficiency. I describe the context of this current study in Chapter 6. Chapter 7 presents the method used in this study, followed by the results in Chapter 8. The last four chapters (Chapters 8-12) are discussion, implications, limitations, and conclusion, respectively.

## **CHAPTER 2 CHARACTERISTICS OF ENGLISH LEARNERS (ELs)**

ELs are a fast-growing segment of the K-12 student population in the United States. The number of EL students enrolled in grades K-12 increased by more than 1,000,000 from 2000-2001 to 2017-2018 (Office of English Language Acquisition, 2021). As of 2017-2018, over 5 million ELs were enrolled in U.S. public schools, accounting for 10% of the total K-12 student population (Hussar et al., 2020).

While national and state statistics often describe ELs as predominantly Hispanic or Latino, Spanish-speaking, U.S.-born, economically disadvantaged, and academically low-achieving, such a general description masks important heterogeneity and instability within the EL population. A key defining characteristic of EL students is their demographic diversity (National Academies of Science, Engineering, Medicine, 2017). For example, ELs in U.S. public schools speak over 400 different languages at home. The top five non-English home languages spoken by ELs in 2017-2018 were Spanish (74.8%), Arabic (2.7%), Chinese (2.1%), Vietnamese (1.6%), and Somali (0.8%) (National Center for Education Statistics, 2019). ELs are also members of every major racial/ethnic group (National Academies of Science, Engineering, Medicine, 2017). Hispanic or Latino students comprised the largest proportion (76.4%) of the total EL enrollment in 2017-2018, followed by Asian (10.7%), White (6.6%), and then Black (4.3%) students (Hussar et al., 2020). Although ELs are on average more likely to live in poverty than their non-EL peers, their economic circumstances vary by race/ethnicity: Hispanic ELs are more likely to live in poor families, so are American Indian and Black ELs, whereas White and Asian ELs often grow up in relatively more favorable economic circumstances (National Academies of Science, Engineering, Medicine, 2017). Moreover, ELs include both foreign- and U.S.-born children, attend schools in urbanized and less urbanized areas (suburban, town, rural), and reside in

traditional immigrant destinations (e.g., California and Texas) as well as new destination states (e.g., North Carolina, Georgia, and Pennsylvania; National Academies of Science, Engineering, Medicine, 2017). In sum, the EL population is characterized by substantial variation in home language, race/ethnicity, socioeconomic status, nativity status, geographic distribution, and other demographic attributes.

In addition to being demographically diverse, ELs are a dynamic student population. Compared to other student subgroups defined by race/ethnicity, poverty, gender, and special education status, ELs are a much less stable subgroup (Abedi, 2004). This is because a student's EL status is closely linked to their developing English language skills: Upon school entry, students with low levels of English proficiency are identified as ELs; as these students progress toward proficiency, they are reclassified as former ELs and are no longer part of the EL subgroup. Hence, a given school's EL population is constantly changing as current ELs are reclassified as English proficient and new entrants are identified as ELs (Abedi, 2004). Due to this instability, which defines the EL population itself, members of the EL subgroup are always expected to perform at a lower level on mainstream assessments than their English-proficient peers (Abedi, 2004). The achievement gap between ELs and non-ELs has been consistently highlighted by national and state assessment reports (National Center for Education Statistics, 2017a; National Center for Education Statistics, 2017b; Sugarman & Geary, 2018) and is *expected*. However, as Saunders and Marcelletti (2013) pointed out, these reports may have significantly overestimated the achievement gap between students who are ever classified (even once) as ELs and those who are never classified as ELs by tracking ELs cross-sectionally (i.e., current EL) rather than longitudinally (i.e., ever an EL).

Chapter summary:

- ELs are a highly heterogeneous population in terms of home language, race/ethnicity, socioeconomic status, nativity status, geographic distribution, and other demographic attributes.
- ELs are a dynamic population that is always expected to perform at a lower level on mainstream assessments than their English-proficient peers.

## **CHAPTER 3 LEGAL UNDERPINNINGS**

### **3.1 Current Education Policies Governing the Education of English Learners (ELs)**

K-12 public education in the United States has been partially governed and funded at the federal level by the Elementary and Secondary Education Act (ESEA), and its reauthorizations since 1965. The most recent reauthorization of ESEA, the Every Student Succeeds Act (ESSA), Pub. L. No. 114-95, 129 Stat. 1802 (2015), was passed in 2015 as a replacement of the influential No Child Left Behind (NCLB) Act, Pub. L. 107-100, 115 Stat. 1425 (2002), which was authorized from 2002 through 2015. What distinguishes NCLB and ESSA from their predecessors is their increased requirement for accountability in having obtained federal funding, and a strong reliance on large-scale, high-stakes testing as part of the accountability system.

To meet ESEA's goal of advancing educational equity, NCLB required states and local educational agencies to develop and implement test-based accountability systems for all students, including those in four historically underachieving subgroups: racial or ethnic minority (i.e., African Americans, Latinos, Asian/Pacific Islanders, Native Americans); economic disadvantaged; special education; and limited English proficient (LEP). One particular subgroup that NCLB has drawn attention to is what the law formerly called LEPs. An LEP was defined as a school-aged individual whose difficulties in reading, writing, understanding, or speaking English deny the person (a) the ability to attain proficiency on a state standardized test, (b) the ability to perform adequately in classrooms when English is the language of instruction, and (c) the opportunity to fully participate in society. 20 U.S.C. § 7601(25) (2014), amended by Every Student Succeeds Act (ESSA), Pub. L. No. 114-95, 129 Stat. 1802 (2015). Without changing this definition, ESSA replaced LEP, a term carrying a negative connotation (e.g., Wiley & Wright,

2004, p. 154), with a more neutral term, English learner (EL; August & Hakuta, 1998).

Following ESSA, I refer to those students as ELs in this manuscript.

NCLB held states, districts, and schools accountable for ELs' academic achievement in two ways: EL performance on standardized content assessments determined the allocation of Title I funds, and their progress toward attaining the state English language proficiency (ELP) standards determined both Title I and Title III funds (Winke, 2011). Title I of NCLB required states to bring all students, including ELs, to the state-defined proficient level in reading (or English language arts) and math by 2014, as measured by performance on state standardized tests. Schools and districts were monitored to progress toward this goal through a mechanism known as Adequate Yearly Progress (AYP). In order for a school to meet the AYP targets in a given year, the school must demonstrate that the state-established annual measurable objectives (AMOs) were achieved by all its students and by ELs and all other student subgroups. In addition to Title I AMOs, schools and districts serving ELs were further held accountable by two Title III annual measurable achievement objectives (AMAOs), which were based on ELs' ELP test performance. Under Title III, schools had to demonstrate that they were achieving (a) annual increases in the number or percentage of ELs making progress in learning English; and (b) annual increases in the number or percentage of ELs attaining English proficiency. Schools that failed to meet the AYP goals faced increasingly severe consequences, which at times eventually included school closures or the loss of federal funds. By establishing these requirements, NCLB aimed at closing achievement gaps for ELs and other student subgroups that had been previously overlooked (Chubb, Linn, Haycock, & Wiener, 2005).

Although ESSA no longer requires states to set AMOs or AMAOs, ESSA maintains many of NCLB's accountability mandates related to ELs (Burke, Morita-Mullaney, & Singh, 2016). In



addition, ESSA moves accountability for performance on ELP assessments from Title III, the title directly addressing EL education, to Title I, the first and largest federal education title, to ensure successful English acquisition and equitable educational opportunities for ELs (Brooks, 2016; for more information about ESSA, see <https://www2.ed.gov/policy/elsec/leg/essa/legislation/title-i.html>). Under ESSA, schools and districts must continue to demonstrate that ELs (a) are making annual progress as measured by standardized tests of ELP and academic content, and (b) are, as a student subgroup, achieving the same level of proficiency on state standards as their non-EL peers (i.e., monolingual English-speaking students, students from non-English-speaking homes who are fluent in English at the time of school entry [initial fluent English proficient], and former ELs who progress out of the EL status [reclassified as full English proficient]; Abedi, 2008). What is different under ESSA is that states now have greater flexibility in determining ELP and academic achievement targets for ELs (i.e., ESSA no longer requires states to meet the universal goal that every student in every school be proficient in reading and math within a pre-determined time frame).

Although NCLB and ESSA are praiseworthy for increasing national awareness about ELs, there is evidence that the laws' mandates for test-based accountability systems have resulted in substantial negative consequences on ELs and schools that serve them (Menken, 2006; 2008). In particular, critics argued that NCLB functioned as a de facto language policy for ELs (Menken, 2008). For example, NCLB shifted focus from helping ELs develop bilingual skills to helping ELs develop their English proficiency *only* by replacing Title VII, the *Bilingual Education Act* (BEA), which had been part of ESEA since 1968, with Title III, the *English Language Acquisition Act* (ELAA) (Haas, 2014; Wiley & Wright, 2004). A second example of NCLB as language policy can be seen in NCLB's accountability requirement for high-stakes testing in

English, which discouraged schools from offering bilingual instruction to ELs out of concern that learning two languages may adversely affect ELs' acquisition of English and thus negatively impact their performance on ELP and mainstream standardized tests: The assumption was that such effects would bring negative consequences to the schools (Wiley & Wright, 2004). Due to these reasons, the number of bilingual programs has considerably decreased across the country following the passage of NCLB, particularly in states where anti-bilingual education policies were already in place prior to NCLB (e.g., Menken & Solorza, 2014; Slama, 2014).

### **3.2 EL Identification, Classification, and Reclassification**

Federal civil rights legislation and federal case law entitle ELs to specialized instructional services that support both their English language development and content proficiency attainment (Linguanti & Cook, 2013). Specifically, Title VI of the Civil Rights Act of 1964 (Public Law 88-352) declares that

“No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance.” (42 USC Sec.2000d)

Title VI requirements were upheld by the U.S. Supreme Court's 1974 *Lau & Nichols* ruling, which stated that rather than providing ELs with the same instructional services (facilities, textbooks, teachers, and curriculum), schools and district must “take affirmative steps” to teach English to ELs in order to provide them with meaningful education and equal access to academic content. Following the *Lau* decision, congress incorporated and extended its principles into the Equal Educational Opportunities Act (EEOA) of 1974 (Public Law 93-380). Under EEOA, states

must ensure that an education agency “take[s] appropriate action to overcome language barriers that impede participation by its students in its instructional programs” (20 USC Sec.1703(f)).

Currently, ESSA provides schools with Title I and Title III funding for the specialized instructional services that ELs are entitled to by law. In exchange for federal funding, ESSA requires schools to (a) identify and classify ELs who need language support, (b) provide ELs with services that support their English language development and content proficiency attainment, and (c) determine when ELs have reached proficiency and can be reclassified as English proficient. An EL is defined under ESSA as a school-aged individual whose difficulties in reading, writing, understanding, or speaking English deny the person (a) the ability of attain proficiency on a state standardized test, (b) the ability to perform adequately in classrooms when English is the language of instruction, and (c) the opportunity to fully participate in society (20 U.S.C. § 7601(25)). This definition requires states to collect at least two sources of information to identify, classify, and reclassify ELs: (a) students’ language minority status, and (b) their level of English proficiency (Abedi, 2008). Because of a lack of consensus among researchers and policymakers as to what it means to be proficient in English as a second language (e.g., Cummins, 2008, Hakuta, Butler, & Witt, 2000), individual states have to, on a practical level, establish their own definition of proficiency. As described in detail below, states (and sometimes districts within a state) have historically varied in their EL identification, classification, and reclassification procedures, with consensus growing only in recent years as to how ELP should be defined and assessed (Liquanti, Cook, Bailey, & MacDonald, 2016).

### ***3.2.1 EL Initial Identification***

Most states (all but four states as of 2013) use a home language survey for initial EL identification (Liquanti & Cook, 2013). The survey is usually administered to parents when

they enroll their child into the school district. If the parents indicate that a language other than, or in addition to, English is spoken at home, then their child is identified as a potential EL and must take an ELP assessment for EL classification.

In their review, Bailey and Kelly (2012) noted several validity issues with states' home language survey practices, including having an unclear or ambiguous survey purpose, using locally-variable, sometimes construct-irrelevant, survey questions, and employing inconsistent administration and decision-making procedures. Linquanti and Cook (2013) further pointed out that the issues identified by Bailey and Kelly (2012) may result in two types of EL identification errors: "false positives" (students wrongly identified as potential ELs when they are in fact not) and "false negatives" (students not properly identified as ELs because they are omitted from initial assessment). Given these findings, the U.S. Department of Education released research-based guidelines for EL identification as part of a large EL tool kit in 2015 in order to guide state and local educational agencies through the process of home language survey development and administration (U.S. Department of Education, & Office of English Language Acquisition, 2017).

### ***3.2.2 EL Classification***

Classification starts once the home language survey identifies a student as a potential EL. States have historically varied in how they define and assess the ELP construct due to lack of federal guidance (Kieffer & Parker, 2016; Thompson, 2017). For example, Wolf and colleagues (2008) found that states used 30 different ELP assessments during the 2006-2007 school year. In recent years, most states have joined two ELP consortia and have adopted common ELP standards and assessments within each consortium, although several states (e.g., Arizona, California, New York, Texas) still develop their own ELP standards as non-consortia, "stand-

alone” educational agencies (Linguanti & Cook, 2013). The larger consortium, known as the World-Class Instructional Design and Assessment (WIDA), now has 40 members, including 35 states, the District of Columbia (DC), one federal agency, and three U.S. territories (<https://wida.wisc.edu/memberships/consortium>). The other consortium, English Language Proficiency Assessment for the 21<sup>st</sup> Century (ELPA21; <https://www.elpa21.org/contact-us/>), is currently composed of 7 states (Arkansas, Iowa, Nebraska, Ohio, Oregon, Washington, and West Virginia). To facilitate comparison of EL outcomes across consortia and states, researchers, representatives, and stakeholders from the two ELP consortia and several stand-alone states have recently developed reference performance level descriptors that map state/consortium-specific ELP levels to a common proficiency scale (Cook & MacDonald, 2014).

### ***3.2.3 EL Reclassification***

EL students are required to take the state ELP assessment to demonstrate English language development in reading, writing, listening, and speaking every year until they meet reclassification requirements. Once reclassified as full English proficient (FEP), the EL is no longer eligible for specialized language services and is moved to the mainstream classroom. Reclassification is therefore a high-stakes decision for individual ELs. As Linguanti and Cook (2013) summarized based on empirical research, “a premature exit may place a student who still has linguistic needs at risk of academic failure, while unnecessarily prolongation of EL status (particularly at the secondary level) can limit educational opportunities, lower teacher expectations, and demoralize students” (p. 20).

In light of the significant consequences of reclassification, individual states are allowed to establish their own reclassification criteria based on the state’s specific context but are required to use, at the minimum, an ELP assessment to do so (Greenberg Motamedi, Singh, & Thompson,

2016). Linquanti and Cook (2015) observed that during the 2015-2016 school year, 29 states and the District of Columbia relied solely on the state ELP assessment for reclassifying ELs, whereas 21 states used the ELP assessment and up to three additional reclassification criteria, which included academic content test results ( $n = 17$ ), local educator input or evaluation (e.g., course grades, scored writing samples;  $n = 15$ ), and input from parents or other stakeholders ( $n = 6$ ). The researchers additionally observed variation in the types of ELP assessment scores (the overall composite score and domain scores) and the associated cut scores that states used to benchmark “English proficient.”

Despite the variation in reclassification criteria across states, Linquanti and Cook (2015) determined that these patterns represented “a notable consolidation of the number and kind of reclassification criteria used” compared to previous years (p. 89). In particular, they noted that the number of states that used a state ELP assessment as the exclusive criterion for exit increased from 12 in 2006-2007 (as reported by Wolf et al., 2008) to 29 in 2015-2016. Such a trend of states increasingly relying solely on the state ELP assessment for reclassification decisions, along with the trend of them joining large ELP assessment consortia in recent years, suggests that states are moving toward a more consistent definition of ELs and English proficiency over time.

#### Chapter summary:

- Current accountability systems require schools and states to monitor ELs’ progress toward English proficiency by assessing those students annually in reading, writing, speaking, and listening using standardized tests.
- In exchange for federal funding, schools must identify and classify ELs who need language support, provide ELs with services that support their English language

development and content proficiency attainment, and determine when ELs have reached proficiency and can be reclassified as English proficient.

- States (and sometimes districts within a state) have historically varied in their EL identification, classification, and reclassification procedures. A consensus has only emerged in recent years as to how ELP should be defined and assessed.

## CHAPTER 4 TIME TO ENGLISH PROFICIENCY

Helping ELs develop sufficient English proficiency to succeed in the mainstream classroom is of great interest to educators, policymakers, and researchers. A large number of studies have attempted to understand the typical length of time that ELs need to attain proficiency in the school context. These studies can be roughly classified into two types depending on whether they were conducted before or after the passage of the 2002 No Child Left Behind (NCLB) Act. Prior to NCLB's assessment and accountability mandates, researchers were limited in their ability to investigate ELs' progress toward English proficiency because high-quality, longitudinal datasets were either nonexistent or rare (Burke, Morita-Mullaney, & Singh, 2016). Using primarily cross-sectional data, pre-NCLB researchers reported a range of times required by EL children to become proficient in English: four to eight years (Collier, 1987), five to seven years (Cummins, 1981), and four to seven years (Hakuta, Butler, & Witt, 2000). While these early studies advanced the understanding of ELs' English acquisition patterns, the estimates these researchers computed could be misleading because the effect of time is confounded with the effect of cohort in cross-sectional designs (Little, 2013).

Student-level longitudinal data have accumulated rapidly since NCLB required states to *annually* assess ELs in reading, listening, speaking, and writing until they became reclassified as proficient in English. This spawned a growing number of studies that explored trajectories of ELs' English language development from a longitudinal perspective. The majority of post-NCLB studies have focused on patterns of EL reclassification (Burke et al., 2016; Conger, Hatch, McKinney, Atwell, & Lamb, 2012; Greenberg Motamedi, Northwest, & Thompson, 2016; Kieffer & Parker, 2016; Slama, 2014; Thompson, 2017; Umansky & Reardon, 2014). As shown in Table 1, researchers in states such as California (Thompson, 2017; Umansky & Reardon,



2014), Florida (Conger et al., 2012), Massachusetts (Slama, 2014), New York (Conger et al., 2012; Kieffer & Parker, 2016), and Washington (Greenberg Motamedi et al., 2016) used their state's (or district's) definition of reclassification to estimate the time it takes for ELs to become reclassified as former ELs. Although those researchers employed a common statistical method known as discrete-time survival analysis, they obtained considerably different time-to-reclassification estimates. In general, researchers in Florida, Massachusetts, New York, and Washington found that half of the ELs were reclassified within three to four years of school entry (Conger et al., 2012; Kieffer & Parker, 2016; Greenberg Motamedi et al., 2016; Slama, 2014), whereas researchers in California reported longer median times to reclassification of up to eight years (Thompson, 2017; Umansky & Reardon, 2014). Variation was also observed in estimates across districts within the same state. For example, Umansky and Reardon's (2014) analysis of the data from one California district yielded considerably lower reclassification rates than Thompson's (2017) analysis of a different California district.

Table 1. Summary of post-NCLB, longitudinal studies estimating the time that ELs take to become reclassified/attain proficiency

Study	State	Student Sample(s)	Data Span (Years)	English language proficiency (ELP) Assessment(s)	Primary outcome of interest	Reclassification criteria	Findings
Conger (2009)	New York	33,706 ELs ages 5-10 who entered New York City public schools between 1996-1999 (report based on the 1997 student cohort only, N = 8976)	1996-2004 (8 years)	Language Assessment Battery (LAB) from 1996-97-2001/02; New York State English as a Second Language Achievement Test (NYSESLAT) in 2002/03	ELP attainment	a) Scoring above the 40th percentile on the LAB from 1996/97-2001/02; b) Attaining proficient level on the NYSESLAT in 2002/03	a) 50% of ELs reach proficiency within 3 years after entry; b) Time that students needed to become proficient increased the older they were when they entered the school system.

Table 1 cont'd

Uman- sky & Rear- don (2014)	California	5,423 Latino ELs (9 cohorts) who entered a large, urban California district in kindergarten between 2000- 2008	2000-2012 (12 years)	California English Language Development Test (CELDT)	Reclassi- fication	a) Attaining an overall CELDT score of 4 or 5 (out of 5) with no subscore (reading, listening, speaking, writing) lower than 3; b) Having a Mid- Basic score on the English Language Arts section of the California Standards Test (CST-ELA); c) Teacher approval; d) Reaching a minimal threshold for GPA in middle and high school	a) The median time to reclassification for Latino kindergarten entrants is 8 years; b) Through 5th grade, English proficiency was a larger barrier for students than was the academic ELA criterion, beginning in 6th grade, academic criterion became a larger barrier to reclassification; c) English immersion students have more favorable outcomes (reclassification rate, ELP attainment, content proficiency attainment) in elementary grades while students in two- language programs catch up and surpass their English immersion peers in middle school.
--	------------	--	-------------------------	--	-----------------------	---	---

Table 1 cont'd

Slama (2014)	Massachusetts	5,354 ELs (one cohort) who entered Massachusetts public schools in kindergarten in 2002-2003	2002-2009 (7 years)	Not specified by the author	Reclassification	Not specified by the author	a) The median time to reclassification is 3 years; b) Spanish-speaking, low-income ELs took longer to become reclassified; c) More than half of the reclassified students scored below proficient on statewide content-area assessments.
Burke, Morita-Mullaney, & Singh (2016)	Indiana	4,010 ELs (one cohort) who entered Indiana public schools in third grade in 2008-2009	2008-2013 (5 years)	LAS Links	ELP attainment as a proxy for reclassification	Attaining Level 5 (out of 5) on LAS Links	Spanish home language status, low SES, and special education status are negatively associated with the odds to reclassification.

Table 1 cont'd

Thompson (2017)	California	202,931 students (8 cohorts) who entered a large California district in kindergarten between 2001- 2008	2001-2010 (9 years)	CELDT	Reclassi- fication	a) Attaining an overall score of 4 or 5 (out of 5) on the CELDT, as well as 3 or higher in the domains of reading, listening, speaking, and writing; b) Scoring at the Basic level or above on the CST-ELA test; c) Teacher evaluation, parent input, and peer comparisons	a) Slightly more than half of ELs who entered in kindergarten were reclassified after six years; b) Boys, native speakers of Spanish, students with lower levels of initial academic L1 proficiency, students in special education, and students whose parents had lower levels of education all have lower probabilities of reclassification than their peers; c) There appeared to be a reclassification window during the upper elementary grades, and by this time students not reclassified became less likely ever to do so, with content-area assessments representing the largest barrier to reclassification for them.
--------------------	------------	--	------------------------	-------	-----------------------	---	--

Table 1 cont'd

Greenberg, Mota-medi, Singh, & Thompson (2016)	Washington	16,957 ELs (7 cohorts) who entered seven poorest and lowest-achieving districts in kindergarten between 2005-2011	2005-2013 (8 years)	Washington Language Proficiency Test II (2005/06-2011/12); Washington English Language Proficiency Assessment (2012/13-2014/15)	Reclassification	Achieving the transitional level (highest of the four levels) on a state ELP assessment	a) The median time to reclassification for ELs who began in kindergarten was 3.8 years; b) Female, Chinese-, Vietnamese-, or Russian/Ukrainian-speaking ELs with advanced initial English proficiency were more likely to be reclassified than male, Spanish- or Somali-speaking ELs with basic/intermediate levels of initial proficiency.
--	------------	---	---------------------	---	------------------	---	---

Table 1 cont'd

Slama et al. (2017)	Texas	71,140 Hispanic ELs (one cohort) who entered Texas public schools in first grade in 2005- 2006	2005-2013 (8 years)	Texas English Language Proficiency Assessment System (TELPAS)	ELP attain- ment	a) Meeting proficiency benchmarks on the TELPAS or other state-approved ELP tests in listening, speaking, and writing; b) Scoring proficient on norm-referenced standardized tests of reading and writing; c) Having a satisfactory teacher evaluation	a) Around 50% of the students who were not yet English proficient by entry to grade 2 attained English proficiency (as measured by the TESLPAS) within 2.6 years, and about 88% were proficient by the end of grade 8; b) Students who began grade 1 with a beginner level of EP, those who participated in a special education program, those who started grade 1 at age 7 or older, and those eligible for the federal lunch program were less likely to attain English proficiency at any grade.
---------------------------	-------	--	------------------------	--	------------------------	---	---

Table 1 cont'd

Conger, Hatch, McKinney, Atwell, & Lamb (2012)	New York and Florida	9,108 ELs ages 5-10 who entered New York City public schools in 1997-1998; 12,158 ELs ages 5-10 who entered Miami-Dade county public schools in 2003-2004	New York: 1997-2003 (6 years); Florida: 2003-2008 (5 years)	New York: LAB; Florida: Miami-Dade County Oral Language Proficiency Scale-Revised (M-DCOLPS-R)	Reclassification	Not specified	a) The median times to reclassification in New York City and Miami were 2.72 and 2.36 years, respectively; b) In both school districts, students who were eligible for free/reduced-priced lunch program, Hispanic, black, and older upon entry have longer median times to proficiency than other students.
Kieffer & Parker (2016)	New York	229,249 ELs (7 cohorts) who entered New York City public schools in kindergarten through 7th grade between 2003/04-2010/11 (with the exception of the 2008/09 cohort)	2003-2012 (8 years)	NYSESLAT; LAB for initial EL classification	Reclassification	Attaining proficient level on the NYSESLAT	a) Slightly more than half of students who entered K as ELs were reclassified within 4 years; b) The median time to reclassification was significantly longer for ELs who entered school in a higher grade, who had lower levels of English proficiency at school entry, and who had specific learning disabilities or speech or language impairments.



Table 1 cont'd

Kim, Curby, & Winsler (2014)	Miami	18,532 low-income, Spanish-speaking ELs (4 cohorts) who participated in the Miami School Readiness Project and who were in kindergarten through fifth grade between 2003-2004 through 2008-2009	2003-2009	M-DCOLPS-R between 2003 and 2007, and Comprehensive English Learner Assessment (CELLA) between 2006-2009	ELP attainment	Between 2003 and 2007: Attaining proficient level (Level 5) on the M-DCOLPS-R; Between 2006 and 2009: Attaining proficient level (Level 5) on the CELLA (although the test changed, the standards for proficiency did not)	a) It took about 2 years for half of the sample to become proficient in English.
------------------------------	-------	---	-----------	--	----------------	--	--

The differences in estimates of time to reclassification across post-NCLB studies are likely related to these studies' two common limitations. First, most of these studies used data from nearly a decade ago, a period when states struggled to define and measure ELP in a valid and consistent way to meet the requirement of NCLB (Abedi, 2008). At the start of NCLB implementation, states had to use different existing, often flawed, ELP assessments to identify and reclassify ELs while waiting for new assessments to become available (Abedi, 2008; Wolf et al., 2008). Wolf et al. (2008), for example, reported that 30 different ELP assessments were used across the country during the 2006-2007 school year. It is only in recent years that there has been growing consensus among states as to how ELP should be defined and assessed. Particularly, most states have joined two large assessment consortia known as WIDA and ELPA 21 (as explained in Chapter 3, Section 3.2.2), with WIDA currently having 40 members including the District of Columbia. States within the same consortium share common ELP assessments (including placement tests and annual ELP assessments) and standards, although some variation still exists within consortia in how states interpret the standards and set cut scores for benchmarking different proficiency levels (Cook and MacDonald, 2015).

Second, post-NCLB researchers' time-to-reclassification estimates are limited by their state's (or district's) specific reclassification criteria, methods, and practices, which often include considerations irrelevant to the construct of ELP. Although reclassification is supposed to occur when an EL attains English proficiency, *reclassification* and *ELP attainment* do not necessarily happen at the same time in practice because states and districts often make reclassification decisions on the basis of ELP-irrelevant considerations (Kieffer, Lesaux, & Snow, 2007; Linquanti & Cook, 2015). As observed by Linquanti and Cook (2015), states frequently couple content-area assessments with ELP assessments in their reclassification policies. Whereas states

with shorter time-to-reclassification estimates (e.g., Florida, Massachusetts, New York) tend to rely solely on the state ELP assessment for EL reclassification, states with longer times to reclassification (e.g., California) usually consider additional, content-based criteria (Thompson, 2017). Without a stable, consensus-bound, shared definition of reclassification, at least for research purposes, the outcomes from research on time to reclassification will necessarily be location (state or district) dependent. Likewise, the outcomes will not be generalizable to other states nor districts. Researchers have long argued against the practice of holding ELs to content-based reclassification expectations because those expectations could be challenging even for similarly situated monolingual English-speaking students (Linguanti & Cook, 2015; Umansky et al., 2015). However, some states continue to do so, which may result in unnecessarily prolongation of EL status (Linguanti & Cook, 2015). In addition, the reclassification process is heavily influenced by local practitioners' (e.g., school administrators, teachers) interpretation of the state and federal policies for EL accountability (Estrada & Wang, 2018; Mavrogordato & White, 2017). In many cases, reclassification may misrepresent proficiency attainment because school administrators prioritize some accountability-related considerations over ELs' proficiency achievements. For example, Kieffer et al., (2007) described the following irony regarding reclassification under the NCLB: On the one hand, an EL's time to reclassification may be spuriously long if the school intentionally keeps them classified after they become English proficient in order to increase the EL subgroup's average performance to meet the goal of NCLB's Title I; On the other hand, the EL's time to reclassification may be spuriously short if the school prematurely reclassifies them before they become proficient in English under the funding incentive of NCLB's Title III for quickly reclassifying ELs. Moreover, teacher evaluation may trump test-based evidence for reclassification eligibility at the local level. In their

investigation of the reclassification process in Texas, Mavrogordato and White (2017) found that some schools relied almost exclusively on teachers' recommendations in making reclassification decisions even when teachers expressed concerns irrelevant to students' progress in ELP (e.g., introverted personality, lack of leadership skills, disciplinary infractions). Similar findings were reported by Estrada and Wang (2018) for schools in California. In a more recent study, Umansky, Callahan, and Lee (2020) further suggested that teachers' reclassification recommendations tended to be biased toward children from positively stereotyped ethnic or language backgrounds (e.g., Chinese children) and against those from negatively stereotyped backgrounds (e.g., Latinx children). Because of the involvement of ELP-irrelevant considerations, reclassification often fails to serve as a proxy for proficiency attainment. Indeed, some empirical research has highlighted significant discrepancies between ELs who have attained proficiency as measured by ELP assessments and those that are actually reclassified (Umansky & Reardon 2014).

In sum, although post-NCLB studies have significantly improved the understanding of EL reclassification patterns, especially time to reclassification, they have limited ability to inform ELs' English proficiency development because (a) these studies are largely based on outdated, state-dependent definitions and measures of ELP, and because (b) reclassification does not necessarily represent ELP attainment in real-world practice. In view of these limitations, new research is needed that uses more recent, higher quality assessment data to investigate ELs' English proficiency development as measured against ELP-based rather than ELP-irrelevant reclassification criteria.

In this dissertation study, I aim to obtain updated estimates of the time that ELs require to attain English proficiency independent of reclassification considerations. To do so, I use more

recent longitudinal data from a large WIDA state, Michigan. I analyze the time necessary for ELs to attain proficiency on WIDA's ELP assessment, "Assessing Comprehension and Communication in English State-to-State for English Language Learners" (ACCESS for ELLs, ACCESS hereafter). This study is timely because existing EL research has primarily drawn on data from a small number of non-consortia, "stand-alone" states that develop their own ELP standards and assessments, such as California, New York, and Texas (Linguanti & Cook, 2013). Little is known about how ELs progress toward English proficiency in large assessment consortia like WIDA, where states have a more common definition of ELP. Located in a WIDA state, this study also has greater generalizability at the national level. Importantly, estimates of time to proficiency from this study can be compared to similar estimates from the other 39 WIDA states.

However, it should be noted that individual WIDA states tend to have a variety of proficiency requirements for ACCESS. Linguanti and Cook (2015) determined that as of September 2015, some WIDA states (e.g., Maine, Pennsylvania, Utah) used only the overall composite score, whereas other states (e.g., Michigan, Tennessee, Wisconsin) used both the composite score and one or more domain scores. Cut scores for benchmarking "English proficient" on ACCESS, too, vary across WIDA states. For example, the composite cut score for exiting EL status ranged from 4.7 to 6.0 across WIDA states in 2015-2016. Given the variation in ELP requirements across WIDA states, I investigate when students attain *all* as well as *specific* components of Michigan's ELP criteria. This helps improve the comparability and generalizability of this study's estimates within the WIDA consortium. To further ensure the robustness of these estimates, I take into account child background factors that previous research has shown to affect EL children's rate of English acquisition, which I will describe next.

#### Chapter summary:

- Studies examining time to proficiency can be roughly classified into two types depending on whether they were conducted before or after the passage of the 2002 NCLB Act. Pre-NCLB studies were limited in their abilities to investigate time to proficiency due to the lack of high-quality, longitudinal EL data.
- The majority of post-NCLB studies investigated time to reclassification, an event correlated with but not equivalent to proficiency attainment in that reclassification often involves criteria irrelevant to the core construct of ELP. In addition, post-NCLB studies have mostly used data from nearly a decade ago, a period when states struggled to define and measure ELP in a valid and consistent way to meet the requirement of NCLB. The lack of a stable, consensus-bound, shared definition of reclassification limits post-NCLB studies' generalizability, making their estimates necessarily location (state or district) dependent.
- This study aims to provide updated, unbiased estimates of the time that ELs require to attain English proficiency independent from reclassification considerations. To achieve broader generalizability, I situated this study in a large WIDA state, Michigan, so that the estimates from this study can be compared with similar estimates from other WIDA states.

## CHAPTER 5 FACTORS RELATED TO TIME TO ENGLISH PROFICIENCY

The English learner (EL) population in U.S. public schools is enormously diverse. They represent different personal, cultural, linguistic, and educational backgrounds. ELs' differences may affect their language learning processes and thus lead to variable outcomes in their English language development (e.g., Kim, Curby, & Winsler, 2014). In order to effectively help ELs acquire English language proficiency (ELP), it is important that educators and researchers understand how variations among EL subgroups are related to their progress toward proficiency (National Academies of Science, Engineering, and Medicine, 2017).

In this study, I examine subgroups of ELs' time to ELP by drawing on bioecological theories of child development (Bronfenbrenner & Ceci, 1993; Bronfenbrenner & Evans, 2000) and usage-based language acquisition theories (Ellis, O'Donnell, & Römer, 2013; O'Grady, 2008; Tomasello, 2003). According to the bioecological model, child development takes place through a complex system of processes that may be *directly* promoted or hindered by the child's personal characteristics and the immediate settings of home and school where the child lives and studies. At the same time, the child may be *indirectly* affected by the more remote contexts of societal attitudes, ideologies, and cultures through the immediate contexts (Bronfenbrenner & Ceci, 1993; Bronfenbrenner & Evans, 2000). Focusing specifically on language development, usage-based theories similarly assert that a child's language acquisition is driven by factors both internal and external to the child. As part of a family of emergentist theories, usage-based approaches assume that children and adult both rely on domain-general cognitive processes in language learning (Behrens, 2009). Prior research has shown that child-internal factors like language aptitude and cognitive maturity (usually measured by chronological age) are significantly related to the rates at which children acquire a second language (Dörnyei & Skehan,

2003; Paradis, 2011). In addition, usage-based theories emphasize the essential role of environmental factors (e.g., home-, school-, and community-related factors) in child language development, asserting that environment provides the raw material, or input, out of which children construct linguistic inventories (Tomasello, 2003). Researchers predict that input properties (quality and quantity) would play a prominent role in explaining individual differences among children in rates and outcomes of language acquisition (Behrens, 2009; O’Grady, 2008).

### **5.1 Overview of This Study’s Six Factors of Interest**

Guided by these theoretical approaches, I investigate how time to proficiency varies by six independent factors of interest. Before defining these factors in detail, below I first provide readers an overview of the six factors:

Two of the independent factors pertain to ELs’ personal characteristics:

1. Primary disability type: whether the EL is identified with a particular type of primary disability;
2. Primary home language: the language that the EL speaks or hears at home.

Two pertain to the home context:

3. Home English use: whether the EL uses or is exposed to English at home;
4. Poverty status: whether the EL lives in an economically disadvantaged home (as indicated by the EL’s eligibility for free/reduced-price lunch at school).

Two pertain to the school context:

5. Instructional programming: whether the EL participated in a particular English-language-learning program at school;
6. Retention: whether the EL is retained in grade.



I chose to examine the six factors mentioned above because EL subgroups defined by these factors have received insufficient attention from educators, researchers, or policymakers despite the need to appropriately serve these EL subgroups. As the recent review by the National Academies of Science, Engineering, and Medicine (2017) suggested, existing studies on ELs' English language development are exclusively or largely based on samples of low-income Hispanic Spanish-speaking ELs, with most studies including only ELs who do not have disabilities. The bias in the existing research's sampling procedure has left many important EL subgroups understudied (e.g., ELs who speak Arabic, Chinese, or other frequently reported non-Spanish home languages, ELs with disabilities). Moreover, researchers of previous studies that did examine those six factors tended to focus on only one or two of them, often times operationalizing these factors in an inadequate manner (e.g., treating poverty status as a time-invariant variable). More research is therefore needed to develop a better understanding of the relationship between those factors and ELs' time to proficiency. In what follows, I describe the six factors of interest and previous studies on those factors as they relate to this study in greater detail.

## **5.2 A Closer Look at This Study's Six Factors of Interest**

### ***5.2.1 Primary Disability Type***

Under the Individuals with Disabilities Education Act (IDEA), states and districts are responsible for locating, identifying, and evaluating all students, including ELs who have a disability that impedes their academic and overall development and are in need of special education and related services (Office of English Language Acquisition, 2020). In 2018-2019, around 14% of ELs ( $n = 714,400$ ) were identified with one or more of 13 disability categories listed in IDEA. The two most common categories among disabled ELs were specific learning

disability (SLD; 48.0%) and speech of language impairment (SLI; 18.1%) (Office of English Language Acquisition, 2020; for definitions of SLD and SLI, see National Academies of Science, Engineering, and Medicine, 2017, p. 354-357). While one would expect disabilities to have negative effects on ELs' English language development, the extent to which this is true is unclear for different types of disabilities (Kieffer & Parker, 2016). Several studies that treated disabled ELs as a homogenous group reported that ELs with disabilities were significantly less likely to become reclassified than ELs without disabilities (Thompson, 2017; Umansky, Thompson, & Díaz, 2017). Challenging the usefulness of this general observation, Kieffer and Parker (2016) uncovered substantial heterogeneity within the disabled EL population. Specifically, they found that the median time to reclassification for ELs with SLDs was about four years longer than that of their peers without disabilities, whereas the median time to reclassification for ELs with SLIs was two years longer than that of ELs without disabilities, suggesting that these two subgroups of disabled ELs may need different types of interventions.

Although the Kieffer and Parker's (2016) findings are informative, the findings provided limited insights into the effects of different disabilities on ELs' English acquisition process because reclassification often involves ELP-irrelevant considerations that may unfairly disadvantage ELs with disabilities, such as content-area assessments and teacher recommendation. In this study I aim to address this limitation by investigating the relationship between disability type and ELs' time to English proficiency. Following Kieffer and Parker (2016), I focus on the two most common disabilities categories among ELs, SLD and SLI, given that the identification of both categories requires language assessments. Although ELs with disabilities are by law entitled to supports for both English language development and learning with a disability, they have been historically underserved at school due to inadequate linguistic

support (Kangas, 2014; 2018) and restricted opportunities to learn (Kangas & Cook, 2020). It is my aim that this study will encourage educators to design better interventions to improve the learning experience for ELs with disabilities by providing information on the English development trajectories of ELs with disabilities in relation to their non-disabled peers.

### ***5.2.2 Primary Home Language***

Although ELs in the United States speak over 400 home languages (Bialik, Scheller, & Walker, 2018), researchers have predominantly focused on the educational outcomes of the largest subgroup, Spanish-speaking ELs, leaving those from other home languages practically unstudied. In time-to-reclassification research, for example, most researchers either exclusively examine reclassification patterns for Spanish-speaking ELs (e.g., Slama et al., 2017; Umansky & Reardon, 2014), or they compare reclassification outcomes between Spanish- and non-Spanish-speaking ELs by treating home language as a dichotomous variable (e.g., Burke, et al., 2016; Slama, 2014). The dichotomous approach to coding home language has led to a general conclusion that non-Spanish-speaking ELs, as a group, are more likely to be reclassified than Spanish-speaking ELs, after controlling for other covariates such as poverty (e.g., Slama, 2014). This conclusion, however, has been challenged by findings from at least two studies where home language was coded more nuancedly (Greenberg Motamedi et al., 2016; Thompson, 2017). Thompson (2017), for example, identified the top five home languages spoken by ELs in a district in California and found that although Cantonese-, Filipino-, and Korean-speaking ELs were about two times more likely to reclassify than Spanish-speaking ELs, Armenian- and Spanish-speaking ELs had similar reclassification rates. These findings highlight the need for a more nuanced understanding of the role of home language on ELs' language acquisition patterns. Such information will allow researchers to better understand distinct patterns of language

development among EL language subgroups and to design interventions to support their language and literacy development (Hammer, Jia, & Uchikoshi, 2011). As Umansky, Callahan, and Lee (2020) observed, the mechanism driving academic differences among EL language and ethnic subgroups may be highly complex and may involve factors at individual, school, and society levels.

### ***5.2.3 Home English Use***

Despite the widely held belief that young children can acquire two languages as quickly and successfully as one, research on language minority children suggests that simply being exposed to two languages does not guarantee successful bilingual development (Hoff, 2018). One important factor that affects the success of language minority children's language acquisition is their language experience, which is often measured by their exposure to and use of the target language (Hoff, 2013; 2018). Language minority children gain language experience in both school and home contexts: while they hear and use the majority and/or minority language for academic or social purposes at school, those children also engage in numerous language activities at home with their parents, siblings, and other relatives.

The role of home language experience on English acquisition has been extensively researched for language minority children in English speaking countries (Bohman, Bedore, Peña, Mendez-Perez, & Gillam, 2010; De Cat, 2020; Duursma et al., 2007; Gutiérrez-Clellen & Kreiter, 2003; Hammer, Komaroff, Rodriguez, Lopez, Scarpino, & Goldenstein, 2012; Hammer, Davison, Lawrence, & Miccio, 2009; Hwang, Mancilla-Martinez, Flores, & McClain, 2019; Jia & Aaronson, 2003; Mancilla-Martinez & Lesaux, 2011; Paradis, 2011; Paradis, Rusk, Duncan, & Govindarajan, 2017; Paradis, Soto-Corominas, Chen, & Gottardo, 2020; Quiroz, Snow, & Zhao, 2010; Reese, Garnier, & Gallimore, 2000). These studies have highlighted substantial variability

in language minority children's English experience at home, including variation in the quality and quantity of English exposure, the access to English literacy resources and activities, and the amount of English use with different interlocutors (Duursma et al., 2007; Hammer et al., 2012; Hwang et al., 2019; Hammer et al., 2009). Despite such variation, researchers have found a strong trend of language shift in the home from using the home language (L1) to English after children start school (Hammer et al., 2009; Kim, Barron, Sinclair, & Jang, 2020), and such a shift often occurs earlier for younger arrivals than for older arrivals (Jia & Aaronson, 2003). Parental beliefs about dual language development may also affect the likelihood and timing of language shift in the home (Hwang et al., 2019).

Regarding the effect of home English experience on language minority children's English achievement and growth, prior research has shown mixed results. The majority of studies examining prekindergarten or kindergarten children's home English experience and English proficiency have found that the more children hear and use English at home with different interlocutors, the more likely they are to demonstrate higher English proficiency (e.g., Bohman et al., 2010; De Cat, 2020; Quiroz et al., 2010; see Hammer et al., 2009 for an important exception). These findings lend support to the time-on-task hypothesis, which predicts a direct relationship between exposure in a language and the likelihood of proficiency in that language (Mancilla-Martinez & Lesaux, 2011). These findings, however, are not always replicated among school-aged children. For example, in their investigation of second-grade Spanish-speaking students, Gutiérrez-Clellen and Kreiter (2003) observed that the amount of English input at home did not impact the children's English language outcomes. Duursma et al. (2007) reported similar findings for fifth-grade Latino ELs who spoke Spanish as their home language. Duursma et al. attributed their findings to the abundant and rich English input that ELs were exposed to at

school, arguing that English exposure at school is so powerful that parental use of English at home is not required for EL children to become or stay proficient in English (see Pearson, 2007 for a discussion of the “critical mass” input hypothesis). An alternative explanation for the lack of relationship between home English exposure and English proficiency is that language minority parents do not have adequate English proficiency to support their child’s English development (Hoff, 2013; Paradis, 2011; Paradis et al., 2017), but this hypothesis has not been tested empirically. One should note that most of the studies reviewed so far adopted a cross-sectional design and only observed language minority children at a single time point (with Hammer et al., 2009 being an important exception). Such a design limits those studies’ abilities to examine a potentially changing relationship between home English experience and English proficiency over time, which may have given rise to their mixed findings. One of the rare longitudinal studies, for example, suggests that the effect of home English exposure on English proficiency may be strongest in prekindergarten through kindergarten years and may decrease rapidly over time (Mancilla-Martinez & Lesaux, 2011). More longitudinal research that captures the co-varying patterns between home English use and English proficiency is needed to better understand the relationship between language minority children’s home English experience and their English acquisition. It is particularly important to focus on EL children, a subgroup of language minority children who are still in the process of learning English at school.

#### ***5.2.4 Poverty Status***

In the United States, EL status is frequently confounded with low socioeconomic status (SES; Capps, Fix, Murray, Ost, Passel, & Herwantoro, 2005; Goldenberg, Rueda, & August, 2006). Existing research has consistently documented the negative impact of poverty on ELs’ linguistic and academic outcomes. Upon entry to kindergarten, ELs from low-income homes

have been found to lag behind their non-economically disadvantaged EL and non-EL peers in English language and literacy skills, and such gaps usually remain large through middle and high school years (Burke et al., 2016; Kieffer, 2008; 2010; 2011; Kim et al., 2014; Slama, 2014; Slama et al., 2017; Thompson, 2017). In their six-year longitudinal study of ELs in Miami, Kim et al. (2014) found that ELs from low-SES homes, as proxied by their eligibility for free/reduced-price lunch programs, showed lower levels of initial English proficiency in kindergarten (as measured by Miami's state ELP assessment), and showed lower odds of meeting Miami's state benchmark for English proficiency compared to those not in poverty. Similar findings have been reported by researchers who investigate factors relating to ELs' time to reclassification (Burke et al., 2016; Slama, 2014; Thompson, 2017). In general, those researchers conclude that ELs who ever received free/reduced-price lunch take significantly longer times to become reclassified than those who were never eligible for subsidized meals. Compared to ELs from middle and high-income homes, ELs from low-income homes are found to have less access to high-quality language exposure, literacy resources, and language-learning activities in English, which may partially explain the SES-related achievement gaps observed by researchers (Gathercole, 2002; Hoff, 2003; Luo et al., 2021).

While informative, the existing literature is limited in describing the dynamic relationship between long-term poverty patterns and English proficiency development. This is because researchers have predominantly treated poverty as time-invariant (i.e., ever eligible for free/reduced-price lunch = 1; otherwise = 0) in their longitudinal studies. In real life, a child's poverty status may change from year to year depending on their household income, thus ELs can have different, long-term poverty patterns. Indeed, Micheltore and Dynarski (2017) identified three distinct poverty groups in their investigation of eighth-grade ELs in Michigan: always in

poverty, occasionally in poverty, and never in poverty. They found that while half of the ELs were ever eligible for a subsidized meal from kindergarten through eighth grade, only 14% had been eligible for meal subsidies in every grade since kindergarten. In addition, their findings indicated that children who were always eligible for a subsidized meal (14%) scored on math tests 0.94 standard deviations below those who were never eligible for meal subsidies and 0.23 below those who were occasionally eligible. These findings thus cast doubt on treating poverty as time-invariant in longitudinal research. In this study, I provide a robust understanding of the dynamic relationship between ELs' changing poverty status on their time to proficiency. I did so by operationalizing poverty as a time-varying variable.

#### ***5.2.5 Instructional Programming***

While there is wide consensus among policymakers, educators, and researchers that ELs should receive specialized language services at school in order to acquire the English proficiency that they need to succeed in mainstream classrooms, the debate has been fierce as to what type of program best serves such a purpose. Controversy has centered particularly on the language of instruction, that is, whether ELs should be instructed only in English or bilingually by drawing on both English and the EL's home language. In the 1970s and 1980s, favorable policies and practices supported bilingual education, in which children were taught partially in their L1, and then the education transitioned to English-only instruction at some point during the elementary grades (Slavin, Madden, Calderón, Chamberlain, & Hennessy, 2011). However, bilingual programs have been declining across the nation since the 1990s due to the influence of the English-only movement (Cheung & Slavin, 2012), which is a type of public policy that has been adopted or enacted in several states. For example, California, Arizona, and Massachusetts have enacted English-only language education policies that generally curtail bilingual education in



public schools in pursuit of rapid English acquisition and reclassification among ELs (Slavin et al., 2011; Umansky & Reardon, 2014). In alignment with growing English-only policy-adoption nation-wide, the federal government passed the No Child Left Behind (NCLB) Act in 2002 to hold states and schools accountable for ELs' performance on standardized content assessments in English and for ELs' progress toward English proficiency as measured by standardized ELP assessments (see more about the requirements of NCLB in Chapter 3, Section 1). These test-based accountability requirements have further discouraged the use of bilingual education in the last 20 years (Cheng & Slavin, 2012). As I explain in Chapter 3, Section 1, the successor of NCLB, Every Student Succeeds Act (ESSA), continues to rely on English-only high-stakes testing as part of the accountability system.

The theories underlying these English-only education policies invoke two general claims: first, exposing ELs only to English will help them develop English language and literacy more quickly (also known as the time-on-task hypothesis; Rossell & Baker, 1996); second, time spent learning the L1 will, in effect, slow down ELs' acquisition of English (Cummins, 2019). Proponents of bilingual education, however, argue that ELs acquire literacy concepts more easily and successfully in their L1 while developing English proficiency. Once proficient in their L1, bilingual education proponents suggest, ELs will be able to transfer those skills and knowledge into English (e.g., Cummins, 1979; Goldenberg, 2008; see MacSwan, Thompson, Rolstad, McAlister, Lobo, 2017 for an overview of two related theoretical positions). Bilingual advocates also argue that a bilingual classroom provides a more friendly and welcoming learning environment for ELs because it is more likely to value cultural and linguistic diversity than a monolingual classroom (Gándara & Orfield, 2010). A large body of research comparing literacy outcomes between ELs instructed in English-only programs and those instructed bilingually has

found either no differences in outcomes measured in English or that ELs in bilingual programs outperform those instructed only in English on English literacy outcomes (see August & Shanahan, 2006; National Academies of Science, Engineering, and Medicine, 2017; Slavin & Cheung, 2005 for syntheses of research; and see McField & McField, 2014; Rolstad, Mahoney, & Glass, 2005 for meta-analyses). However, there is a dearth of research that has explored if ELs in bilingual programs also perform equal to or better than their peers in English-only programs in acquiring English language skills.

One important exception to the dearth of research on the effect of instructional programming on ELs' English proficiency outcomes is Umansky and Reardon's (2014) work. These two researchers compared time to reclassification among Latino ELs in four language-learning programs in one district in California: English immersion; transitional bilingual; maintenance bilingual; and dual immersion bilingual programs. Before describing Umansky and Reardon's study in more detail, I first briefly define these four distinct language-learning programs for readers:

- a. English immersion: Designed to teach English and content knowledge to ELs in an accessible manner. ELs share their classroom with mainstream, non-EL students, and instruction is solely in English.
- b. Transitional bilingual: Designed to develop proficiency in English and content knowledge through L1 instruction. All students in the transitional bilingual speak Spanish as their L1. Instruction is predominantly given in Spanish in kindergarten (80%-90%) with increasing proportions of English in each subsequent grade. The program ends after the third grade at which point students transfer to an all-English environment.

- c. Maintenance bilingual: Designed to develop proficiency in both Spanish and English. All students in the transitional bilingual speak Spanish as their L1. Instruction is 80% to 90% Spanish in kindergarten, transitioning to 50% to 65% English by the fifth grade. The maintenance bilingual program continues throughout elementary school, and in some cases throughout middle and high school.
- d. Dual immersion bilingual: Designed to serve both ELs and English-speaking students in the same classroom. The goal is for both groups of students to become bilingual in Spanish and English. Elementary school instruction shifts from predominantly in Spanish in kindergarten to half Spanish and half English by the end of elementary.

Following nine cohorts of 5423 Latino ELs from kindergarten through grade 11 over years 2000-2001 through 2011-2012, Umansky and Reardon found that ELs enrolled in bilingual programs attained English proficiency at a slower pace in elementary school but had higher overall English proficiency outcomes (including reclassification, English literacy proficiency, and English reading, speaking, listening, and writing proficiency) than English-immersion ELs by the end of high school. These findings suggest that bilingual programs are associated with significant benefits on English proficiency attainments and that such benefits may manifest in the medium to long term until in late elementary school and beyond.

Although Umansky and Reardon have contributed significantly to the understanding of the role of instructional programming on ELs' English acquisition patterns, their findings may have limited generalizability because the study was located in a single district in California, a stand-alone state that is not a member of any of the two large ELP assessment consortia in the United States (i.e., WIDA and ELPA 21, see Chapter 3, Section 3.2.2 for a description of these two consortia). Building upon Umansky and Reardon, this current study tests if the findings about

bilingual advantage can be generalized to the WIDA consortium by using data from a large WIDA state, Michigan (see Chapter 6 for a detailed description of the context of Michigan).

### **5.2.6 Retention**

In the United States, around 9% to 11% of students in kindergarten through eighth grade are retained in grade every year (Planty et al., 2009). The rate of retention varies by a number of student characteristics, with males, ethnic minorities, ELs, and students who live in economically disadvantaged homes on average having a higher risk of being retained (Tavassolie & Winsler, 2019; Jimerson & Renshaw, 2012). According to the U.S. Department of Education (2016), ELs are overrepresented among retained students in every grade except kindergarten, with the percentage of ELs retained being about twice the percentage of ELs enrolled in each grade in high school (9<sup>th</sup> through 12<sup>th</sup> grades). Students in U.S. public schools can be held back for two reasons: On the one hand, students may be recommended for retention by teachers, school administrators, or parents using traditional policies (based on student classroom performance and behavior), sometimes regardless of their academic performance; On the other hand, they may be retained due to retention policies that some states and school districts have implemented (e.g., Texas, Florida, New York, Chicago) based on high-stakes tests (Tavassolie & Winsler, 2019). Test-based retention policies usually require students to repeat a particular grade (e.g., third grade) if they perform below grade-level standards on content-area assessments (e.g., a state-approved reading assessment), with retention acting as a remediation tool for low-achieving students.

Much research has been devoted to investigating the effects of retention on the long- and short-term outcomes for students (e.g., Jimerson, 1999; Jimerson & Ferguson, 2007; Jacob & Lefgren, 2004; Green & Winters, 2007; Ou & Reynolds, 2010; Schwert, West, & Winters, 2017;

Winters & Greene, 2012;). Studies examining traditional retention suggest that although retention leads to short-term gains in academic subjects (e.g., reading, math), it harms student performance in the long run (see Allen, Chen, Wilson, & Hughes, 2009; Holmes, 1989; Jimerson, 2001 for meta-analyses) and increases dropout rates in high school (see Jimerson, Anderson, & Whipple, 2002 for a research synthesis). The inefficacy of retention has been confirmed by a recent international meta-analysis (Valbuena, Mediavilla, Choi, & Gil, 2020). However, research that focuses on test-based retention in the United States tends to suggest less negative and sometimes positive retention effects on student performance (Jacob & Lefgren, 2004; Roderick & Nagaoka, 2005; Schwert et al., 2017; Winters & Greene, 2012). For example, Schwert et al. (2017) found that Florida's third-grade reading retention policy resulted in positive effects for retained students over time, with those students continuing to outperform their promoted peers in the same grade through grade eight in math and grade ten in reading. Also located in Florida, Figlio and Ozek (2020) reported a number of positive effects for retained ELs relative to their promoted EL peers, including reduced time to proficiency, decreased likelihood of taking a remedial English course in middle school, doubled likelihood of taking an advanced course in math and science in middle school, and tripled likelihood of taking credit-bearing courses in high school. However, those findings may not generalize to other contexts given that Florida requires all retained children to attend summer school. Because those researchers were not able to disentangle the effects of retention from the effects of instructional support, it is unknown whether and to what extent the observed positive effect is attributable to retention itself.

Despite the large body of research on retention, few studies focus on ELs, a population more susceptible to negative effects of retention than non-ELs. Particularly, little is known about

the effect of retention on ELs' language and literacy development. This study takes an initial step in understanding this issue by investigating the relationship between traditional retention on ELs' time to proficiency. This investigation is timely because many states, including Michigan, are implementing test-based retention policies with only limited exemption criteria for ELs (usually for newcomers). Understanding the number of ELs retained based on traditional policies and the effect of traditional retention for ELs over time can help policymakers make more informative judgments about the effectiveness of test-based retention in improving ELs' language proficiency.

Chapter summary:

- Drawing on bioecological theories of child development (Bronfenbrenner & Ceci, 1993; Bronfenbrenner & Evans, 2000) and usage-based language acquisition theories (Ellis, O'Donnell, & Römer, 2013; O'Grady, 2008; Tomasello, 2003), I investigate the effects of the following independent factors on ELs' time to proficiency: primary disability type, primary home language, home English use, poverty status, instructional programming, and retention.
- Prior research has suggested that disabilities would negatively ELs' likelihood of reclassification; however, it remains unclear to what extent different types of disabilities may slow down ELs' English language development.
- Existing studies have concluded that Spanish ELs, as a group, are less likely to attain proficiency than their peers who do not speak Spanish at home. More research is needed to test if this general conclusion applies to all non-Spanish-speaking EL subgroups.

- Studies examining the relationship between home English use and English proficiency have shown mixed results, perhaps because such a relationship may change in direction and magnitude over time. More longitudinal research is required to determine whether and to what extent home English use contribute to ELs' English proficiency development.
- ELs who have ever lived in poverty are known to reclassify at a slower rate than their peers who have never lived in poverty. However, researchers have yet to take account of the time-varying nature of poverty status in longitudinal research.
- The vast research examining the effect of instructional programming on ELs' literacy development has either found no differences between ELs instructed bilingually and those instructed only in English in English literacy outcomes or that ELs in bilingual programs outperform their peers in English-only programs. Researchers still need to explore whether this observation extends to the domain of English proficiency development.
- Prior research has yielded mixed results with regard to the effect of retention on the academic outcomes of low-achieving students. More research needs to be conducted to investigate the effect of retention on ELs and their short-term and long-term English proficiency outcomes.

## CHAPTER 6 CONTEXT FOR THE STUDY

This study was located in Michigan, a state with a moderate and growing immigrant population. Although the growth rate of foreign-born population in Michigan has slowed between 2000 and 2016 (27%) as compared to the rate between 1990 and 2000 (47%), it far outpaces the growth rate of native-born population (Sugarman & Geary, 2018). According to the Michigan Department of Education (MDE), children who are English learners (ELs) in the public schools account for around 6% of the state K-12 student population, compared to 10% nationwide. Nearly two-thirds of Michigan's school-aged ELs are born in the United States, and around 87% of them have only native-born parents (Sugarman & Geary, 2018). The five top home languages spoken by Michigan's ELs are Spanish, Arabic, Bengali, Chinese, and Albanian (Sugarman & Geary, 2018). In 2015-2016, around 40% of Michigan's ELs spoke Spanish at home (Sugarman & Geary, 2018), which was considerably lower than the proportion of Spanish-speaking ELs in most states and at the national level (75%). While Arabic is spoken by 2% of ELs nationwide, almost a quarter of ELs in Michigan self-reported to have Arabic as their home language in 2015-2016 (Sugarman & Geary, 2018).

By definition, and as explained in Chapter 2, ELs are not expected to perform on mainstream assessments at a level comparable to that of their English-proficient peers. In Michigan, considerable achievement gaps have been observed between ELs and non-ELs in various academic domains (e.g., reading, math, science, and social studies), with the gap growing larger at older grade levels (Sugarman & Geary, 2018). Despite these gaps, graduation rates among Michigan's ELs have been increasing over the last five years. As of 2015-2016, the proportion of ELs to graduate within four years was 69%, compared to a four-year graduation



rate of 80% for all students (Sugarman & Geary, 2018). These rates are comparable to those nationwide.

### **6.1 EL Identification, Classification, and Reclassification Practices in Michigan**

Michigan joined the WIDA consortium in 2013. Since then, it has adopted the WIDA English language development (ELD) Standards (WIDA, 2020) as its state ELP Standards and WIDA's ELP assessment, ACCESS for ELLs (hereafter ACCESS), as its state ELP assessment. During the time of the study, the MDE required schools to initially identify students as ELs through a home language survey at school entry. Students who spoke a language other than English at home were subsequently screened for EL status using the WIDA ACCESS Placement Test (W-APT) and a state-approved early literacy or reading test. Those who scored below a designated level on the W-APT or below grade level in reading were classified as ELs. Once classified, ELs were placed in specialized English-language-learning programs and were required to take ACCESS every spring until they met the requirements for reclassification. To be reclassified, ELs had to score proficient on ACCESS and at grade level on state-approved reading and writing assessments. The requirement for literacy proficiency was dropped in fall 2020 (MDE, 2020).

In fall 2013, EL experts in the MDE set the cut scores for benchmarking English proficiency as measured by ACCESS. The MDE re-evaluated the appropriateness of the cut scores on a yearly basis and made several adjustments during the years of data collection, i.e., 2013-2014 through 2018-2019. The cut scores for defining "English proficient" on ACCESS between 2013-2014 and 2018-2019 are as follows:

- 2013-2014 through 2014-2015: A minimum overall composite score of 5.0 (out of 6.0) on ACCESS, and a minimum score of 5.0 (out of 6.0) in the four domains of speaking, listening, reading, and writing
- 2015-2016: A minimum overall composite score of 5.0 (out of 6.0), and a minimum score of 4.5 (out of 6.0) in the four domains of speaking, listening, reading, and writing
- 2016-17 through 2018-2019: A minimum overall composite score of 4.5 (out of 6.0), and a minimum score of 4.0 (out of 6.0) in reading and writing

The most recent adjustment was made in fall 2020, when the MDE removed the requirements for domain scores and changed the cut score for the overall composite score to 4.8 (out of 6.0). For this study, I define English proficiency in alignment with Michigan’s minimum requirements for being proficient on ACCESS for the years of data collection. An EL was considered “English proficient” in a particular year if the student scored at or above the minimum required scores for the overall composite score and domain scores on ACCESS in that year.

Some of the adjustments that the MDE made to the proficiency cut scores for ACCESS were in response to the changes that WIDA introduced to the test and interpretation of test scores. In summer 2016, WIDA conducted standard-setting studies (which, in educational measurement, are formal processes to establish performance standards or benchmark scores) to re-examine ACCESS proficiency level scores. These studies were motivated by the following factors: migrating from a paper-and-pencil to an online assessment, employing a new centrally scored, revised speaking assessment, and adapting to the influence of college and career ready standards (Cook & MacGregor, n.d.). Although no changes were made to the design of ACCESS in the year of 2016-2017, the proficiency standards were updated and purposefully rendered more challenging. As a result, ELs would have to showcase higher language skills to achieve the

same proficiency level scores in and after 2016-2017 than before that year. According to WIDA, the scale scores before and after the 2016 standard-setting are not comparable (email communication with WIDA's Client Relations Specialist, Andrea Traverse, 2019). Neither is a common scale for linking the two sets of scores available (email communication with WIDA's Client Relations Specialist, Andrea Traverse, 2019). It is important to note that despite those changes, the ordinal, six-level system of WIDA's ELD standards and its interpretation stayed the same.

## **6.2 EL Instructional Programs in Michigan**

During the time of the study, Michigan public schools provided various language-learning programs for ELs using two broad approaches: the English as a second language (ESL) approach, in which English is the primary language used for instruction, and the bilingual approach, in which English and the student's home language are both used for instruction (National Academies of Sciences, Engineering, and Medicine, 2017). Those instructional programs drew from 10 identified program models, including five ESL models (i.e., Sheltered Instruction Observation Protocol [SIOP], ESL Instruction, Sheltered ESL Instruction, Structured English Immersion, and Content-based ESL), four bilingual models (i.e., Bilingual Dual-Language Instruction, Bilingual Two-Way Immersion, Transitional Bilingual Instruction, and Bilingual Heritage Language Instruction), and the newcomer model (which could be guided by either the ESL or bilingual approach). General descriptions of these models can be found below. Like in many other states and districts (e.g., California, Indiana, Massachusetts), ESL models are more widely used in Michigan public schools than other program models (see Thompson, 2017 for data in California; see Burke et al., 2016 for data in Indiana; see Slama, 2014 for data in Massachusetts). The most popular ESL model during the study period was ESL Instruction,

followed by Sheltered ESL Instruction, and Structured English Immersion. Among bilingual program models, Transitional Bilingual Instruction was most frequently employed, followed by Dual-Language Instruction.

Because neither the codebook nor the websites of MDE and CEPI include information about how each program model is defined, operationalized, and implemented, I provide general descriptions for some those programs based on a consensus study report by the National Academies of Sciences, Engineering, and Medicine (2017). This report distinguishes among three ESL models: the ESL model, the content-based ESL model, and the sheltered instruction model, and two bilingual models: transitional bilingual model and the dual-language model. Next, I describe each model in detail based on the report.

**ESL model.** In ESL instructional programs, ESL-certified teachers provide explicit language instruction that focuses on the development of proficiency in English skills (e.g., grammar, vocabulary, and communication). Students may have a dedicated ESL class in their school day or may receive pull-out ESL instruction, wherein they work with a specialist for short periods during other classes.

**Content-based model.** In content-based programs, ESL-certified teachers use content as a medium for building language skills. Although using content as a means, the instruction is still focused primarily on learning English. Similar to traditional ESL programs, the content-based program may be taught during a dedicated ESL class or pull-out ESL instruction.

**Sheltered Instruction model.** In sheltered instruction programs, instruction is focused on the teaching of academic content rather than the English language itself. Teachers (general education or ESL-certified) use specialized techniques (e.g., visual aids, physical activities) to

make academic instruction in English understandable to ELs. Sheltered instruction typically takes place in EL-only classrooms and is designed specifically for ELs.

**Transitional bilingual model.** In transitional bilingual programs, students typically begin learning in the home language in kindergarten or grade 1 and transition to English incrementally over time. Students in early-exit programs typically exit prior to grade 3, whereas those in late-exit programs may exit as late as grade 5. Although the L1 is used to leverage English, the goal of transitional bilingual programs is to achieve English proficiency as quickly as possible. The division of the languages across instructional times and content areas may vary from program to program.

**Dual-language model.** Unlike transitional bilingual programs, which are oriented towards English proficiency, dual-language programs aim to help students develop high levels of language and literacy proficiency in English and a partner language. Dual-language programs can be further classified into two types based on the students they serve. One-way dual-language programs primarily serve students from the same home language group (e.g., ELs acquiring English and developing their L1), whereas two-way dual-language programs, also known as two-way immersion programs, serve both ELs and their English-speaking peers by providing instruction in English and the ELs' L1 (or the partner language). A two-way immersion classroom is usually comprised of half native English speakers and half native speakers of the partner language. Unlike some ESL programs where quick EL reclassification is the goal, the dual-language program is designed to keep ELs enrolled through its completion regardless of when and whether they attain proficiency in English. Heritage program is one type of dual-language program specifically designed for heritage language learners studying English and their heritage language (e.g., Navajo).

In this study, I investigated the role of instructional programming on the time it takes ELs to attain ELP by comparing ESL and bilingual instruction. Students in different ESL programs were treated as a homogenous subgroup. Newcomer programs and other unidentified programs were also considered ESL programs given that the language of instruction was undetermined for those programs. As such, the results should be more strictly interpreted as a comparison between non-bilingual versus bilingual instruction, where non-bilingual instruction was predominantly delivered via an ESL model. Within bilingual models, I further differentiated between transitional programs and dual-language programs. Based on the descriptions above, I labeled programs using Transitional Bilingual Instruction *transitional bilingual programs* and those using Bilingual Dual-Language Instruction, Bilingual Two-Way Immersion, and Bilingual Heritage Language Instruction *dual-language bilingual programs*. Students who received no language instruction constituted another subgroup in this comparison. Table 2 displays how I classified and coded different program models.

Table 2. Classification and coding of EL instructional programs provided by Michigan public schools

EL instructional program in Michigan public schools	Classification/coding	Classification/coding justification
1. Sheltered Instruction Observation Protocol (SIOP)	ESL	The instruction language of the seven program models is English only. (Newcomer programs and unidentified programs were also considered ESL programs because the language of instruction was undetermined for those programs.)
2. ESL Instruction		
3. Sheltered ESL Instruction		
4. Structured English Immersion		
5. Content-based ESL		
6. Newcomer		
7. Unidentified EL program model		

Table 2 cont'd

8. Transitional Bilingual Instruction	Transitional bilingual	Instruction is provided in both the EL's home language and English. Students typically begin learning in the home language in kindergarten or grade 1 and transition to English incrementally over time. This category includes both early- and late-exit transitional bilingual programs.
9. Bilingual Dual-Language Instruction	Dual-language	Instruction is provided in both the EL's home language and English. Unlike transitional bilingual programs, which are oriented towards English proficiency, dual-language programs aim to help students develop high levels of language and literacy proficiency in English and a partner language. This category includes both one-way and two-way dual language programs.
10. Bilingual Two-Way Immersion	bilingual	
11. Bilingual Heritage Language Instruction		
12. No program	No program	Students in this category receive no specialized language services (e.g., parents opt out of the services).

## Chapter summary:

- This study was located in Michigan, a large WIDA state. In 2013, Michigan joined the WIDA consortium and adopted WIDA's ELD Standards as the state ELP Standards and WIDA's ELP assessment, ACCESS, as the state ELP assessment.
- The MDE established proficiency criteria for ACCESS in 2013 and adjusted those criteria several times over the period of this study (2013-2014 through 2018-2019).
- Michigan public schools offer a variety of program models to ELs. In this study, I focus on the differential effects of ESL, transitional bilingual, dual-language bilingual, and no program on ELs' time to proficiency.

## CHAPTER 7 RESEARCH QUESTIONS AND METHOD

### 7.1 Research Questions

With this study, I estimated the average time it took English learners (ELs) in Michigan public schools to attain English proficiency as measured against the state English language proficiency (ELP) standards. The study was guided by the following research questions:

- 1) How much time does it take ELs in Michigan public schools to attain English proficiency?
- 2) What is the relationship between time to proficiency and the following factors that may affect EL children's rate of English acquisition?
  - a. Primary disability type
  - b. Primary home language
  - c. Home English use
  - d. Poverty status
  - e. Instructional programming
  - f. Retention
- 3) How much time does it take ELs in Michigan public schools to first attain each individual ELP criteria for ACCESS for ELLs (i.e., Michigan's state ELP assessment, hereafter ACCESS), including
  - a. the minimum composite score?
  - b. the minimum score in reading?
  - c. the minimum score in writing?
  - d. the minimum score in speaking?
  - e. the minimum score in listening?



## 7.2 Method

### 7.2.1 Data

For this study, I requested from the Michigan Educational Research Institute (MERI; <https://miedresearch.org>) six years of administrative data on six cohorts of students who entered Michigan public schools in kindergarten as ELs between 2013-2014 and 2018-2019. Each student was followed for one to six years depending on the year that they entered kindergarten. The earliest cohort was followed for six years (from 2013-2014 through 2018-2019), and the latest cohort for one year (2018-2019). Two datasets containing annual student-level test and demographic data were provided by the Michigan Department of Education (MDE) and the Center of Educational Performance and Information (CEPI) in Michigan, respectively, via the MERI. I relied on a unique student identifier and a school-year indicator to link the two data files. Data on students are available for as long as they remained classified as ELs in Michigan public schools during the study period of 2013-2014 through 2018-2019. I focused on this six-year period because it captures the complete data between Michigan's first ACCESS administration (spring 2014) and its last uninterrupted administration in 2019 before COVID-19 hit (in March 2020, the Michigan ACCESS testing window was truncated due to COVID-19 and the pandemic's related state-wide public-school closures).

I cleaned the test and demographic datasets separately and then merged them to obtain the analytic dataset for this study. How I prepared the data for this study is described in detail in Appendix A *Data Preparation*. The final analytic sample consisted of 54,146 students (161,211 records) who were identified as ELs upon kindergarten entry. Table 3 displays demographic characteristics of the full analytic sample separately by cohort (2013-2014 kindergarten entrants are referred to as K13-14 cohort, 2014-2015 kindergarten entrants are referred to as K14-15

cohort, and so on). The sample was gender-balanced, consisting of primarily Hispanic or Latino (35.71%), White (36.77%), and Asian (23.28%) racial and ethnic groups. The majority the students were from Spanish- or Arabic-speaking, low-income homes (as measured by eligibility for free-/reduced-price lunch). About 10% of the students were ever in a transitional or dual-language bilingual program, and a similar percentage of students were ever in a special-education program due to specific learning disabilities, speech and language impairment, or other types of disabilities. Several characteristics in Table 3 are time-varying, meaning that their values may change from year to year. Below, I provide more information on those time-varying characteristics, focusing particularly on whether and how their values changed over time.

Table 3. Mean characteristics of the full analytic sample by cohort

Student characteristic	Full sample ( <i>n</i> = 54,146)		K13-14 ( <i>n</i> = 9,566)		K14-15 ( <i>n</i> = 9,049)		K15-16 ( <i>n</i> = 9,429)		K16-17 ( <i>n</i> = 9,126)		K17-18 ( <i>n</i> = 8,649)		K18-19 ( <i>n</i> = 8,327)	
Gender	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Female	26040	48%	4573	48%	4482	50%	4583	49%	4355	48%	4108	47%	3939	47%
Male	28106	52%	4993	52%	4567	50%	4846	51%	4771	52%	4541	53%	4388	53%
Race														
Hispanic or Latino	19333	36%	3648	38%	3210	35%	3398	36%	3222	35%	3006	35%	2849	34%
African-American or Black	1594	3%	257	3%	233	3%	256	3%	280	3%	294	3%	274	3%
AIAN/NHPI <sup>a</sup>	153	<1%	25	<1%	29	0%	24	<1%	23	<1%	24	<1%	28	<1%
Asian	12603	23%	1985	21%	2096	23%	2297	24%	2066	23%	2076	24%	2083	25%
White	19910	37%	3553	37%	3376	37%	3345	35%	3446	38%	3159	37%	3031	36%
Two or more races	553	1%	98	1%	105	1%	109	1%	89	1%	90	1%	62	1%
Primary disability type														
None	47734	88%	8214	86%	7930	88%	8292	88%	8034	88%	7763	90%	7501	90%
Learning disabilities	949	2%	382	4%	243	3%	189	2%	97	1%	33	0%	5	0%
Speech and language disabilities	3845	7%	687	7%	609	7%	689	7%	711	8%	596	7%	553	7%
Other disabilities	1618	3%	283	3%	267	3%	259	3%	284	3%	257	3%	268	3%
Primary home language (family) <sup>b</sup>														
Spanish	19590	36%	3720	39%	3278	36%	3440	36%	3271	36%	3008	35%	2873	35%
Arabic	14062	26%	2338	24%	2353	26%	2404	25%	2450	27%	2295	27%	2222	27%
Bengali	1537	3%	191	2%	218	2%	292	3%	265	3%	288	3%	283	3%
Chinese	1508	3%	282	3%	267	3%	283	3%	233	3%	216	2%	227	3%
Albanian	1061	2%	213	2%	205	2%	163	2%	164	2%	168	2%	148	2%
AfroAsiatic	1775	3%	321	3%	266	3%	242	3%	355	4%	304	4%	287	3%
Austronesian	1363	3%	288	3%	236	3%	251	3%	186	2%	201	2%	201	2%
BaltoSlavic	1419	3%	277	3%	266	3%	260	3%	223	2%	204	2%	189	2%

Table 3 cont'd

Dravidian	2428	4%	318	3%	385	4%	459	5%	418	5%	424	5%	424	5%
IndoIranian	3068	6%	519	5%	554	6%	564	6%	475	5%	482	6%	474	6%
Italic/Germanic	1513	3%	309	3%	251	3%	267	3%	253	3%	239	3%	194	2%
Japanese/Korean	1694	3%	276	3%	283	3%	272	3%	281	3%	276	3%	306	4%
Other	3128	6%	514	5%	487	5%	532	6%	552	6%	544	6%	499	6%
Primary home language														
Spanish	19591	36%	3719	39%	3278	36%	3439	36%	3273	36%	3009	35%	2873	35%
Arabic	14061	26%	2337	24%	2358	26%	2405	26%	2447	27%	2292	27%	2222	27%
Bengali	1539	3%	191	2%	220	2%	292	3%	265	3%	288	3%	283	3%
Chinese	1510	3%	282	3%	267	3%	284	3%	234	3%	216	2%	227	3%
Albanian	1060	2%	212	2%	205	2%	163	2%	164	2%	168	2%	148	2%
Aramaic	1025	2%	124	1%	118	1%	128	1%	249	3%	198	2%	208	2%
Japanese	1030	2%	161	2%	180	2%	158	2%	168	2%	165	2%	198	2%
Telugu	1255	2%	161	2%	210	2%	235	2%	215	2%	215	2%	219	3%
Urdu	860	2%	153	2%	168	2%	143	2%	145	2%	122	1%	129	2%
Vietnamese	971	2%	210	2%	169	2%	176	2%	131	1%	140	2%	145	2%
Other	11244	21%	2016	21%	1876	21%	2006	21%	1835	20%	1836	21%	1675	20%
Poverty														
Never poor	13400	25%	2308	24%	2207	24%	2432	26%	2105	23%	2064	24%	2284	27%
Ever poor	40746	75%	7258	76%	6842	76%	6997	74%	7021	77%	6585	76%	6043	73%
Program <sup>c</sup>														
Ever in transitional bilingual	4028	7%	1888	20%	717	8%	467	5%	421	5%	315	4%	220	3%
Ever in dual-language bilingual	2016	4%	448	5%	352	4%	366	4%	311	3%	305	4%	234	3%
Ever in no program	1101	2%	356	4%	248	3%	194	2%	185	2%	86	1%	32	0%
Only ESL	47122	87%	6945	73%	7759	86%	8414	89%	8214	90%	7949	92%	7841	94%
Home English use														

Table 3 cont'd

Ever using English at home	10834	20%	1681	18%	1630	18%	1867	20%	1919	21%	1970	23%	1767	21%
Never using English at home	43312	80%	7885	82%	7419	82%	7562	80%	7207	79%	6679	77%	6560	79%
Retention <sup>d</sup>														
Ever retained	4333	8%	791	8%	791	9%	920	10%	935	10%	896	10%	NA	NA
Never retained	49813	92%	8775	92%	8258	91%	8509	90%	8191	90%	7753	90%	NA	NA
Ten largest districts														
Dearborn City School	5878	11%	1039	11%	978	11%	996	11%	985	11%	964	11%	916	11%
Detroit Public Schools	2850	5%	470	5%	469	5%	540	6%	414	5%	498	6%	459	6%
Grand Rapids Public Schools	2370	4%	467	5%	390	4%	411	4%	416	5%	358	4%	328	4%
Warren Consolidated Schools	2315	4%	397	4%	344	4%	494	5%	425	5%	369	4%	286	3%
Utica Community Schools	2091	4%	412	4%	386	4%	325	3%	366	4%	318	4%	284	3%
Troy School	2003	4%	334	3%	318	4%	372	4%	316	3%	346	4%	317	4%
Ann Arbor Public Schools	1501	3%	285	3%	305	3%	284	3%	229	3%	209	2%	189	2%
Farmington Public School	1203	2%	213	2%	194	2%	186	2%	198	2%	203	2%	209	3%
Kentwood Public Schools	1184	2%	178	2%	137	2%	247	3%	190	2%	217	3%	215	3%
Plymouth-Canton Community Schools	973	2%	186	2%	202	2%	192	2%	142	2%	128	1%	123	1%

Note. <sup>a</sup>AIAN/NHPI = American Indian or Alaska Native/Native Hawaiian or Other Pacific Islander.

<sup>b</sup>The counts of Spanish, Arabic, Bengali, Chinese, and Albanian speakers differ slightly between primary home language (family) and primary home language coding because some students reported different home languages across the years, and for those students who reported to speak two or more languages equally likely at home, a random language was chosen as their home language.

<sup>c</sup>The percentages of different program categories do not add up to 100% because there was some overlap among students who were ever in transitional bilingual programs, who were ever in dual-language bilingual programs, and who ever had no program.

<sup>d</sup>2487 students were retained once in grade, 15 were retained twice in grade.

### 7.2.1.1 Instructional Program Participation Patterns

It is useful to look at how students in the analytic sample moved through EL instructional programs until they left the dataset. Table 4 presents the 13 most common patterns of EL instructional program participation for the earliest cohort that entered Michigan public schools in kindergarten in 2013-2014 (i.e., K13-14 cohort). As shown in Table 4, the majority of K13-14 students (72.59%) had received only ESL instruction prior to leaving the dataset (regardless of when they left). The most common program pattern for students who were ever in bilingual instruction was one year in a transitional bilingual program followed by one or more years of ESL instruction. A small number of students (less than 2%) remained in transitional or dual-language bilingual programs for six years.

Table 4. The most common EL program participation trajectories for 2013-2014 kindergarten entrants prior to leaving the study ( $n = 9566$ )

EL program pattern prior to leaving the study	Frequency	Percent (out of 9566)
1. 6 years ESL instruction	2718	28.41%
2. 5 years ESL instruction	1500	15.68%
3. 4 years ESL instruction	1058	11.06%
4. 1 year ESL instruction	981	10.26%
5. 1 year transitional bilingual instruction, then 5 years ESL instruction	496	5.19%
6. 2 years ESL instruction	415	4.34%
7. 1 year transitional bilingual instruction, then 4 years of ESL instruction	308	3.22%
8. 3 years ESL instruction	273	2.85%
9. 6 years transitional bilingual instruction	184	1.92%
10. 1 year transitional bilingual instruction, then 3 years ESL instruction	174	1.82%
11. 1 year transitional bilingual instruction	126	1.32%
12. 2 years transitional bilingual instruction, then 4 years ESL program	122	1.28%
13. 6 years dual-language bilingual program	102	1.07%

Tables 5 and 6 display the most common program participation trajectories for K13-14 students who were ever in transitional and dual-language bilingual instruction, respectively. The two tables highlighted an important difference between transitional and dual-language programs: While students in transitional programs typically began bilingual instruction in kindergarten and transitioned to English-only classrooms after one or two years, students in dual-language instruction tended to receive bilingual instruction for a more sustained period of time.

Table 5. The most common program participation trajectories for 2013-2014 kindergarten entrants who ever participated in transitional bilingual programs prior to leaving the study ( $n = 1888$ )

Program pattern	Frequency	Percent (out of 1888)
1 year transitional, then 5 years English-only	496	26.27%
1 year transitional, then 4 years English-only	308	16.31%
6 years transitional	184	9.75%
1 year transitional, then 3 years English-only	174	9.22%
1 year transitional	126	6.67%
2 years transitional, then four years English-only	122	6.46%
5 years transitional	76	4.03%
2 years transitional, then 3 years English-only	40	2.12%
4 years transitional	38	2.01%
1 year transitional, 2 years English-only	36	1.91%

Table 6. The most common program participation trajectories for 2013-2014 kindergarten entrants who ever participated in dual-language programs prior to leaving the study ( $n = 448$ )

Program pattern	Frequency	Percent (out of 448)
6 years dual-language instruction	102	22.77%
5 years dual-language instruction	29	6.47%
1 year dual-language instruction	26	5.80%
5 years dual-language instruction, then 1 year ESL instruction	25	5.58%
1 year ESL, then 1 year dual-language instruction, then 4 years ESL	23	5.13%
4 years dual-language instruction	18	4.02%
1 year dual-language instruction, then 5 years ESL	16	3.57%
2 years dual-language instruction	16	3.57%
1 year ESL, then 5 years dual-language instruction	15	3.35%



It should be noted that bilingual education has been shrinking across the years in Michigan, as shown in Table 7. There was a significant drop in transitional-bilingual enrollments following the year of 2013-2014. The data indicated that this drop happened mainly in the Dearborn City School District, a large school district with the highest percentage of ELs (an average of 47% between 2013 and 2019). Although the reason why this district closed its transitional bilingual programs is unknown, the overall decline in bilingual education in Michigan is likely related to No Child Left Behind (NCLB) Act's accountability requirements, a trend that has also been observed in other states since the implementation of the NCLB (Crawford, 2007; Menken & Solorza, 2014).

Table 7. The numbers and percentages of students in ESL, transitional bilingual, and dual-language bilingual instruction by cohort

Cohort	Only ESL		Ever dual-language bilingual		Ever transitional bilingual	
	Number	Percent	Number	Percent	Number	Percent
K13-14 ( <i>n</i> = 9566)	6945	72.60%	448	4.68%	1888	19.74%
K14-15 ( <i>n</i> = 9049)	7759	85.74%	352	3.89%	717	7.92%
K15-16 ( <i>n</i> = 9429)	8414	89.24%	366	3.88%	467	4.95%
K16-17 ( <i>n</i> = 9126)	8214	90.01%	311	3.41%	421	4.61%
K17-18 ( <i>n</i> = 8649)	7949	91.91%	305	3.53%	315	3.64%
K18-19 ( <i>n</i> = 8327)	7841	94.16%	234	2.81%	220	2.64%

In Michigan, the decline in bilingual education, particularly in transitional bilingual instruction, has disproportionately affected Arabic-speaking ELs. Table 8 displays the numbers and proportions of ELs ever in transitional and dual-language bilingual education in K13-14 and in the entire sample for ten most frequently reported home languages. As Table 8 shows, 46.77% Arabic-speaking ELs in K13-14 were ever enrolled in a transitional bilingual program, compared to 10.93% in the entire sample. This large discrepancy is likely attributable to the close of transitional bilingual programs in the Dearborn City School District, which serves predominantly

Arabic-speaking ELs. Meanwhile, one can see that ELs from different home languages were differentially served by bilingual instruction. Despite the overall low bilingual enrollment, ELs from some home languages (e.g., Bengali, Chinese, Telugu) were substantially less likely to receive instruction in their native language compared to their EL peers from other home languages (e.g., Arabic, Albanian, Japanese, Spanish). Some factors that may have affected the quantity of bilingual programs include the number of ELs, availability of bilingual teachers, textbooks, and teaching materials, and existence of a common dialect shared by ELs from the same home language (Conger, 2010).

Table 8. The numbers and percentages of ELs who were ever in transitional bilingual and dual-language bilingual programs in 2013-2014 kindergarten entrants and in the entire sample by ten most frequently reported home languages

Primary home lang	Program	Number in K13- 14	Percent in K13- 14	Number in full sample	Percent in full sample
Spanish (K13-14 = 3719; full sample = 19591)	Ever in dual-language bilingual	321	8.63%	1717	8.76%
	Ever in transitional bilingual	498	13.39%	1547	7.90%
Albanian (K13-14 = 212; full sample = 1060)	Ever in dual-language bilingual	3	1.42%	5	0.47%
	Ever in transitional bilingual	54	25.47%	188	17.74%
Arabic (K13-14 = 2337; full sample = 14061)	Ever in dual-language bilingual	36	1.54%	89	0.63%
	Ever in transitional bilingual	1093	46.77%	1535	10.92%
Aramaic (K13-14 = 124; full sample = 1025)	Ever in dual-language bilingual	1	0.81%	2	0.20%
	Ever in transitional bilingual	6	4.84%	30	2.93%
Bengali (K13-14 = 191; full sample = 1539)	Ever in dual-language bilingual	0	0.00%	1	0.06%
	Ever in transitional bilingual	5	2.62%	34	2.21%
Chinese (K13-14 = 282; full sample = 1510)	Ever in dual-language bilingual	4	1.42%	8	0.53%
	Ever in transitional bilingual	13	4.61%	34	2.25%

Table 8 cont'd

Japanese (K13-14 = 161; full sample = 1030)	Ever in dual-language bilingual	34	21.12%	99	9.61%
	Ever in transitional bilingual	1	0.62%	4	0.39%
Telugu (K13-14 = 161; full sample = 1255)	Ever in dual-language bilingual	1	0.62%	1	0.08%
	Ever in transitional bilingual	1	0.62%	5	0.40%
Urdu (K13-14 = 153; full sample = 860)	Ever in dual-language bilingual	7	4.58%	12	1.40%
	Ever in transitional bilingual	15	9.80%	49	5.70%
Vietnamese (K13-14 = 210; full sample = 971)	Ever in dual-language bilingual	0	0.00%	6	0.62%
	Ever in transitional bilingual	18	8.57%	47	4.84%
Other (K13-14 = 2016; full sample = 11244)	Ever in dual-language bilingual	41	2.03%	76	0.68%
	Ever in transitional bilingual	184	9.13%	555	4.94%

### 7.2.1.2 Poverty Patterns

Poverty status is measured by the student's eligibility for the federal free/reduced-price lunch program. Table 9 presents the ten most common poverty trajectories for K13-14 students before they left the dataset.

Table 9. The ten most common poverty trajectories in 2013-2014 kindergarten entrants prior to leaving the study ( $n = 9566$ )

Poverty pattern	Frequency	Percent (out of 9566)
6 years in poverty	2861	29.91%
5 years in poverty	1311	13.70%
4 years in poverty	670	7.00%
4 years NOT in poverty	598	6.25%
1 year NOT in poverty	588	6.15%
5 years NOT in poverty	561	5.86%
6 years NOT in poverty	431	4.51%
2 years in poverty	362	3.78%
2 years NOT in poverty	273	2.85%
3 years in poverty	200	2.09%

Around 30% of the students remained in poverty for six years. For those students who left the dataset early (in less than six years), they either never experienced poverty or remained in poverty for as long as they stayed in this study. Although poverty is not a static condition, ELs in this sample tended to have stable family income over the years they were observed.

### 7.2.1.3 Home-English-Use Patterns

Data on home language use was collected by CEPI on a yearly basis. The student or their parents self-reported what language(s) the student spoke at home. CEPI recorded up to five home languages per student per year in the dataset. I was interested in the patterns of home language change in ELs over time, specifically, when ELs started using English at home in replacement of or in addition to their primary home language. Table 10 displays the eleven most common home-English-use trajectories in K13-14 students before they left the study. As one can see, the majority of the students never used English at home during the time they were observed. For those who did use English at home, they likely started using English in their second grade and tended to continue to do so thereafter.

Table 10. The eleven most common home-English-use trajectories in 2013-2014 kindergarten entrants prior to leaving the study ( $n = 9566$ )

Home-English-use pattern	Frequency	Percent (out of 9566)
6 years never using English at home	3219	33.65%
5 years never using English at home	1676	17.52%
4 years never using English at home	1134	11.85%
1 year never using English at home	1089	11.38%
2 years never using English at home	460	4.81%
3 years never using English at home	307	3.21%
2 years no English use, then 4 years using English at home	268	2.80%
2 years no English use, then 2 years using English at home	177	1.85%
2 years no English use, then 3 years using English at home	167	1.75%
4 years no English use, then 2 years using English at home	123	1.29%
6 years using English at home	93	0.97%

#### 7.2.1.4 Retention

Of the full sample, 4333 students were retained in grade at least once during the period of analysis. Among those students, 4317 were retained once in grade, and 16 were retained twice in grade. In the K13-14 cohort, 796 students were ever retained during the 6-year study period, accounting for 8% of all students in that cohort. As shown in Table 11, K13-14 students who ever experienced retention were most likely retained in kindergarten. The number of retentions was increasingly smaller as grade increased.

Table 11. The number of retentions among 2013-2014 kindergarten entrants by grade ( $n = 797$ )

Grade	Number of retentions	Percent (out of 797 retentions)
kindergarten	539	67.63%
grade 1	166	20.83%
grade 2	63	7.90%
grade 3	28	3.51%
grade 4	1	0.13%
grade 5	0	0
grade 6	0	0

*Note.* 785 students were retained once in grade, 6 students were retained twice in grade.

Unfortunately, the data does not show why those students were retained. Descriptive statistics (shown in Table 12) suggest that K13-14 ELs who were ever retained were more likely to be male, disabled, non-White/Asian, and living in low-income families. In addition, retained students in K13-14 achieved lower English proficiency scores (as measured by ACCESS) in each year from school entry (2013-2014) through the last year of data collection (2018-2019). It is unknown, however, if low English proficiency was one of the reasons why those students were retained in grade.

Table 12. Descriptive statistics for 2013-2014 kindergarten entrants by retention status ( $n = 9,566$ )

Categorical variable	Ever Retained		Never Retained	
	$n$	%	$n$	%
Gender				
Female	334	7.30%	4239	92.70%
Male	457	9.15%	4536	90.85%
Primary disability status				
Not disabled	563	6.85%	7651	93.15%
Specific learning disabilities	64	16.75%	318	83.25%
Speech/language impairments	113	16.45%	574	83.55%
Other disabilities	51	18.02%	232	81.98%
Race				
Hispanic or Latino	487	13.35%	3161	86.65%
African-American or Black	24	9.34%	233	90.66%
AIAN/NHPI	2	8.00%	23	92.00%
Asian	90	4.53%	1895	95.47%
White	177	4.98%	3376	95.02%
Two or more races	11	11.22%	87	88.78%
Poverty				
Ever	713	9.82%	6545	90.18%
Never	78	3.38%	2230	96.62%
Continuous variable	$M$	$SD$	$M$	$SD$
ACCESS composite proficiency level				
Year 2013-2014	1.64	0.61	2.94	1.38
Year 2014-2015	3.10	1.21	3.74	0.77
Year 2015-2016	3.77	0.89	4.55	0.93
Year 2016-2017	3.32	0.76	3.89	0.81
Year 2017-2018	3.54	0.76	4.42	0.78
Year 2018-2019	4.06	0.73	4.30	0.74

*Note.* AIAN/NHPI = American Indian or Alaska Native/Native Hawaiian or Other Pacific Islander.

#### 7.2.1.5 Information on School Districts

The sampled students were enrolled in 481 distinct public-school districts across Michigan. Table 13 displays the mean demographic characteristics for the ten school districts with largest EL populations during the period of the study. As shown in Table 13, the districts

varied substantially in the average total enrollment, proportion of economically disadvantaged students, proportion of students in special education, and proportion of ELs, indicating that it is necessary to control for district effects on ELs' time to proficiency.

Table 13. Mean characteristics for the ten districts with largest EL populations between 2013-2014 and 2018-2019

School district	Enrollment	Proportion of economically disadvantaged students	Proportion of students in special education	Proportion of ELs
Ann Arbor Public Schools	17361	23%	11%	8%
Dearborn City School District	20214	71%	8%	47%
Detroit Public Schools Community District	48585	80%	17%	12%
Farmington Public School District	10043	24%	11%	13%
Grand Rapids Public Schools	16467	77%	23%	25%
Kentwood Public Schools	9018	67%	14%	18%
Plymouth-Canton Community Schools	17495	17%	10%	6%
Troy School District	12881	13%	9%	15%
Utica Community Schools	27751	33%	11%	11%
Warren Consolidated Schools	14507	60%	10%	25%

### 7.2.2 Measures

In Michigan, ACCESS is administered to ELs annually in the spring semester to monitor their progress toward proficiency. ACCESS aligns with the World-Class Instructional Design and Assessment (WIDA) English Language Development (ELD) Standards (2020) and assesses ELs' language skills in listening, reading, speaking, and writing along a vertically equated developmental scale at six grade-level clusters: (a) Kindergarten, (b) Grade 1, (c) Grades 2-3, (d) Grades 4-5, (e) Grades 6-8, and (f) Grades 9-12. The entire test is divided into four subsections, each targeting a distinct skill. Table 14 displays the construct measured by each test subsection

(based on Wolf et al., 2008). Multiple-choices questions are used to assess listening and reading, whereas performance-based tasks are used assess speaking and writing. Speech and writing samples elicited from test-takers are scored by certified raters according to the WIDA ELD standards (task-level inter-rater reliability > .70; Center for Applied Linguistics, 2018). Test takers' performances are reported as scale scores (100-600) and proficiency levels (1-6) for each subsection and for combinations of subsections. For example, the overall composite score is a weighted sum of listening (15%), speaking (15%), reading (35%), and writing (35%) scores. A recent technical report from WIDA described satisfactory reliability coefficients (stratified Cronbach's alpha) of 0.90 or above for ACCESS across all grade clusters (Center for Applied Linguistics, 2018). The report also provided rich validity evidence that supported the intended use of ACCESS for the purpose of assessing social and academic English proficiency of EL students.

Table 14. Test constructs of the four ACCESS subsections

ACCESS subsection	Test construct
Listening	Process, understand, interpret, and evaluate spoken language in various situations for 5 content areas: social and instructional, language arts, math, science, and social studies
Speaking	Engage in oral communication in a variety of situations for an array of purposes and audiences for the 5 content areas mentioned above.
Reading	Process, interpret, and evaluate written language, symbols and text with understanding and fluency for the 5 content areas mentioned above.
Writing	Engage in written communication in a variety of forms for an array of purposes and audiences for the 5 content areas mentioned above.

Table 15 displays the full sample's mean ACCESS proficiency levels by skill, cohort, and academic year. The table also shows the number and percentage of students in each cohort that attained proficiency on ACCESS by academic year. In general, the six cohorts demonstrated similar growth patterns in overall proficiency (as measured by the mean composite score) as well as in each of the four domains (reading, writing, speaking, and listening) from the first year (year



1) through the sixth year (year 6) of schooling. However, the “growth” should be interpreted with caution here because for each academic year, only students who had not attained proficiency at the beginning of that year were included in the table (i.e., the sample changed over time). One should also keep in the mind that the proficiency standards for ACCESS were updated and rendered more challenging in 2016-2017 (see more about WIDA’s 2016-2017 standard-setting procedure in Chapter 6). As a result, students would have to obtain higher scale scores to achieve the same proficiency levels in and after 2016-2017 than before that year. Comparing year 1 through year 3’s proficiency-level scores of different cohorts (e.g., K13-14 versus K16-17), one can see that the scores of speaking dropped the most among the four domains after 2016-2017, which indicates that the standards for speaking may have been rendered more difficult than the standards for the other three domains during the standard-setting process.

Table 15. ACCESS proficiency levels for the full sample by skill, cohort, and academic year

Cohort	Academic year	Year since school entry	Number of students (not proficient yet at the beginning of the year)	Attained proficiency by the end of the year		Composite		Reading		Writing		Speaking		Listening	
				<i>n</i>	%	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
K13-14	2013-2014	Year 1	9566	221	2.3%	2.83	1.38	2.80	1.82	2.45	1.16	3.63	1.58	4.45	1.82
	2014-2015	Year 2	8417	17	0.2%	3.68	0.84	4.29	1.20	3.02	0.63	4.38	1.53	4.66	0.90
	2015-2016	Year 3	7918	4	0.1%	4.48	0.95	5.00	1.16	3.28	0.52	4.93	1.44	5.47	0.90
	2016-2017	Year 4	7550	1190	15.8%	3.84	0.83	4.22	1.60	3.62	0.53	3.03	0.72	5.16	1.25
	2017-2018	Year 5	6108	1984	32.5%	4.33	0.82	4.60	1.37	3.90	0.61	3.47	0.75	5.75	0.75
	2018-2019	Year 6	3989	1284	32.2%	4.26	0.74	4.42	1.40	3.96	0.55	3.29	0.77	5.80	0.60
K14-15	2014-2015	Year 1	9049	281	3.1%	2.92	1.43	2.88	1.85	2.52	1.22	3.75	1.59	4.53	1.82
	2015-2016	Year 2	8015	64	0.8%	4.03	1.08	4.50	1.39	3.20	0.70	4.23	1.49	5.01	1.28
	2016-2017	Year 3	7506	630	8.4%	3.62	0.81	4.03	1.52	3.35	0.58	3.03	0.73	5.00	1.35
	2017-2018	Year 4	6557	959	14.6%	3.86	0.78	4.27	1.51	3.69	0.54	3.00	0.69	5.12	1.28
	2018-2019	Year 5	5384	1792	33.3%	4.25	0.80	4.54	1.35	3.90	0.59	3.34	0.77	5.68	0.82
K15-16	2015-2016	Year 1	9429	495	5.3%	2.85	1.49	2.80	1.83	2.52	1.25	3.74	1.56	4.52	1.82
	2016-2017	Year 2	8270	296	3.6%	3.41	0.92	3.81	1.55	2.74	0.66	3.40	0.97	5.18	1.36
	2017-2018	Year 3	7545	661	8.8%	3.68	0.81	4.16	1.46	3.39	0.62	3.00	0.72	5.04	1.37
	2018-2019	Year 4	6541	749	11.5%	3.75	0.75	4.23	1.46	3.59	0.53	2.94	0.71	5.05	1.30
K16-17	2016-2017	Year 1	9126	356	3.9%	2.56	1.19	2.26	1.37	2.15	0.95	3.91	1.85	4.36	1.87
	2017-2018	Year 2	8128	183	2.3%	3.31	0.86	3.78	1.55	2.69	0.64	3.17	1.00	5.19	1.35
	2018-2019	Year 3	7463	429	5.8%	3.56	0.77	4.12	1.41	3.28	0.60	2.91	0.74	4.92	1.39
K17-18	2017-2018	Year 1	8649	332	3.8%	2.57	1.17	2.27	1.37	2.15	0.94	3.96	1.83	4.36	1.86
	2018-2019	Year 2	7674	126	1.6%	3.22	0.81	3.69	1.50	2.61	0.61	3.17	1.10	5.24	1.29
K18-19	2018-2019	Year 1	8327	377	4.5%	2.55	1.19	2.25	1.39	2.12	0.96	3.99	1.82	4.35	1.85

*Note.* In 2016-2017, the proficiency standards for ACCESS were updated and rendered more challenging. ELs would have to showcase higher language skills (i.e., obtain higher scale scores) to achieve the same proficiency level scores in and after 2016-2017 than before that year. However, it is important to note that the ordinal six-level system of WIDA's English language development standards and its interpretation stayed the same.

### 7.2.3 Variables

A summary of outcome variables, predictors, and control covariates can be found in Table 16.

#### 7.2.3.1 Outcome Variables

The outcome variables include (a) one primary outcome variable indicating ELP attainment, and (b) three secondary outcome variables representing *attainment of the minimum overall composite, reading, writing, listening, and speaking scores on ACCESS*, respectively. The primary outcome is a dichotomous time-varying variable that I coded 1 if the student met all proficiency requirements for ACCESS (i.e., reaching the minimum overall composite and domain scores) in a given year (0 if otherwise). Although the meaning of 1 (attaining proficiency)/0 (failing to attain proficiency) remained the same across the years, the criteria used to determine those values varied according to the MDE's minimum requirements for proficiency on ACCESS (see Chapter 6 for more). Secondary outcome variables were coded similarly. Note that some ELs attained one or more outcomes more than one time during the study period; however, for this study, I only focused on the first event and estimated the time it took ELs to attain a particular outcome for the first time between 2013-2014 and 2018-2019.

#### 7.2.3.2 Predictors

In what follows, I briefly describe the predictors of this study. For specific details about the considerations that I took in coding each predictor, please see Appendix B *Considerations in Coding Predictors*.

**Time in school since kindergarten entry.** Time is a categorical, time-varying variable that records the number of years a student has attended school since entry at kindergarten, regardless of grade repetition. In this study, I defined the beginning of time (time 0) as the

beginning of an EL's first year at school (i.e., the kindergarten year), and the end of time as the year the student attained proficiency or left Michigan public schools (whichever came first). Each year in between was considered a discrete time period. This resulted in a maximum of six time periods (corresponding to the six years in 2013-2014 through 2018-2019) during which the EL was observed. The value of time ranges from 0 to 5, representing time 0 through time 5. For example, for a student in the K14-15 cohort who remained in the study until the last year of data collection, time 0 is 2014-2015, and end of time (time 4) is 2018-2019.

**Primary disability type.** This is a categorical, time-invariant variable that records the primary type of disability that a student was identified with. This variable takes four possible values: no disability, speech and language impairment, specific learning disability, and other disabilities (including all other disabilities). Students with no disabilities were chosen as the reference group.

**Primary home language.** The sampled ELs represented 245 identified home language groups. Because it is difficult to handle this large number of home language groups in statistical analyses, I recoded this variable to obtain a smaller number of categories to represent the students' home language backgrounds. I first identified the five most frequently reported home languages; then, within the remaining languages, I further identified the seven most frequently reported home language families based on the information available from *glottolog* (<https://glottolog.org>). In total, I identified 12 home language categories:

- 1) Top five home language groups: Spanish, Arabic, Bengali, Chinese, Albanian
- 2) Top seven home language family groups:
  - a. AfroAsiatic: Amharic, Aramaic, Oromo, Somali, Syriac, Tigrinya

- b. Austronesian: Filipino, Indonesian, Central Khmer, Malay, Philippine Languages, Tagalog, Vietnamese
- c. Dravidian: Kannada, Malayalam, Tamil, Telugu
- d. BaltoSlavic: Bosnian, Bulgarian, Macedonian, Polish, Russian, Serbian, Ukrainian
- e. IndoIranian: Gujarati, Hindi, Kurdish, Marathi, Nepali, Panjabi, Persian, Pushto, Urdu
- f. Italic/Germanic: French, Italian, Portuguese, Romanian, English
- g. Japanese/Korean: Japanese, Korean

All other languages were coded “other.” Spanish was chosen as the reference category because it was the largest home language group. Primary home language is a categorical, time-invariant variable.

**Home English use.** This is a dichotomous, time-varying variable that indicates whether a student ever self-reported using/hearing English at home in addition to or in replacement of their home language between school entry and a given, later time period (1= Yes; 0 = No). This variable was coded 0 until the first time when the student ever reported using/hearing English at home and was coded 1 thereafter.

**Poverty status.** This is a dichotomous, time-varying variable that indicates whether a student was ever eligible for the federal free/reduced-price lunch program between school entry, and a given, later time period (1 = Yes; 0 = No). This variable was coded 0 until the first time when the student was ever in poverty and was coded 1 thereafter.

**Instructional program.** This is a categorical, time-invariant variable that indicates the type of language-learning program in which an EL was placed when entering school in

kindergarten. This variable takes four possible values: no program (i.e., an EL was not enrolled in any language program), English as a Second Language (ESL) program (including a variety of program models: Sheltered Instruction Observation Protocol, ESL Instruction, Sheltered ESL Instruction, Structured English Immersion, Content-based ESL, newcomer, and other unidentified program models), transitional bilingual program (i.e., Transitional Bilingual Instruction), and dual-language bilingual program (including three dual-language models: Bilingual Dual-Language Instruction, Bilingual Two-Way Immersion, and Bilingual Heritage Language Instruction). The most frequent program type, ESL, was chosen as the reference group.

**Retention.** This is a dichotomous, time-varying variable that indicates whether a student was ever retained between school entry and a given, later time period (1= Yes; 0 = No). This variable was coded 0 until the first time when the student was ever retained in grade and was coded 1 thereafter. Please see Appendix B for some examples of how poverty status, English use at home, and retention were coded.

#### *7.2.3.3 Control Covariates*

Control covariates include gender (dichotomous, time-invariant variable; 1 = female; 0 = male), age (continuous, time-invariant variable measured in years), race/ethnicity (categorical, time-invariant variable with six categories: Hispanic or Latino [reference category], African-American or Black, Asian, White, Two or more races, and a collapsed category composed of Native Hawaiian or Pacific Islander and American Indian or Alaska Native), initial English proficiency in kindergarten (continuous, time-invariant variable ranging from 1 to 6, more explanation about this variable can be found in this chapter, Section 7.2.4), cohort (categorical, time-invariant variable with six categories indicating the year the student entered kindergarten,

K13-14 selected as the reference category), and school district codes (categorical, time-varying variable indicating the district in which a student was enrolled in a given year, the largest school district, Dearborn City School District, selected as the reference category). District codes were included as fixed-effects covariates to control for the characteristics of the school context and unobserved differences in student, family, and community characteristics that may have influenced parents' choice of school (Conger, 2010). The fixed-effects district variables were coded time-varying because some students switched between school in the years observed. In such cases the district code was consistent with the school district the student was enrolled when they were observed.

Table 16. Summary of outcome variables, predictors, and control covariates

Variable	Definition	Operationalization	Coding
<b>Outcome variable</b>			
Proficiency attainment	Whether a student attained all proficiency criteria for ACCESS in a given year	Dichotomous, time-varying	0 = No; 1 = Yes
Composite-score attainment	Whether a student reached the cut score for the composite score in a given year	Dichotomous, time-varying	0 = No; 1 = Yes
Reading-score attainment	Whether a student reached the cut score for the reading domain in a given year	Dichotomous, time-varying	0 = No; 1 = Yes
Writing-score attainment	Whether a student reached the cut score for the writing domain in a given year	Dichotomous, time-varying	0 = No; 1 = Yes
Speaking-score attainment	Whether a student reached the cut score for the speaking domain in a given year	Dichotomous, time-varying	0 = No; 1 = Yes
Listening-score attainment	Whether a student reached the cut score for the listening domain in a given year	Dichotomous, time-varying	0 = No; 1 = Yes



Table 16 cont'd

Predictor			
Time	An indicator defining time periods, with initial time period being the time when a student was enrolled, and the last time period being the last year the student was observed	Categorical	time 0 (initial time period) – time 5 (last time period)
Primary disability type	The type of primary disability a student was identified with	Categorical, time-invariant variable	4 categories: none (reference), specific learning disabilities, speech or language disabilities, other disabilities
Primary home language	The most frequent non-English language (family) a student was identified to use at home	Categorical, time-invariant variable	13 categories: Spanish (reference), Arabic, Bengali, Chinese, Albanian, AfroAsiatic, Austronesian, BaltoSlavic, Dravidian, IndoIranian, Italic/Germanic, Japanese/Korean, Other
Home English use	Whether a student ever self-reported using/hearing English at home in addition to or in replacement of their home language between school entry and a given, later time period	Dichotomous, time-varying variable	0 = No; 1 = Yes
Poverty status	Whether a student was ever eligible for the federal free/reduced-price lunch program between school entry, and a given, later time period	Dichotomous, time-varying variable	0 = No; 1 = Yes
Instructional program	The type of language-learning program in which an EL was placed when entering school in kindergarten	Categorical, time-varying variable	4 categories: ESL (reference), no program, transitional bilingual, dual-language bilingual

Table 16 cont'd

Retention	Whether a student was ever retained between school entry and a given, later time period	Dichotomous, time-varying variable	0 = No; 1 = Yes
Control covariate			
Gender	Whether a student was self-identified as a female or male	Dichotomous, time-invariant variable	0 = Male; 1 = Female
Age	The age at which a student was enrolled in school (measured in years)	Continuous, time-invariant variable	Centered at the mean
Race	The race/ethnicity that a student was identified with	Categorical, time-invariant variable	6 categories: Hispanic or Latino (reference), African-American or Black, Asian, White, Two or more races, collapsed category of Native Hawaiian or Pacific Islander and American Indian or Alaska Native
Initial English proficiency	An EL's end-of-kindergarten ACCESS composite proficiency level	Continuous, time-invariant variable	Centered at the mean
Cohort	The cohort a student belongs to	Categorical, time-invariant variable	
District code	The primary school district a student was enrolled in a given year	Categorical, time-varying variable	481 categories: the Dearborn City School District was the reference category

#### 7.2.4 Analysis

For this study, I conducted two sets of data analyses using different samples. For the first set of analyses, I used the full sample comprised of all six cohorts of EL students (*full sample*); for the second set of analyses, I used a subset of the full sample that consisted of those ELs in the three earlier cohorts (2013-2014 kindergarten cohort, 2014-2015 kindergarten cohort, and 2015-

2016 kindergarten cohort) who had not attained English proficiency by the end of the first year of instruction (*subsample*). I used these two different samples because previous studies suggest that ELs' initial English proficiency upon school entry is confounded with many of this study's focal predictors, particularly instructional programming (Conger, 2010; Slavin, Madden, Calderón, Chamberlain, & Hennessy, 2011, Thompson, 2017; Umansky & Reardon, 2014). In general, those studies show that ELs with high levels of initial proficiency are more likely to be placed in an English only program, whereas those with low levels of initial proficiency are more likely to be placed in a bilingual program if a bilingual program utilizing the EL's first language is available (e.g., Conger, 2010). Given the potential correlation between initial proficiency and program placement, one needs to control for initial proficiency in order to robustly estimate the effect of instructional programming on time to proficiency. This, however, cannot be easily done in this study because in Michigan, although ELs are required to take an English placement test (WIDA screener) when they are first registered in the public school system, schools do not always enter ELs' placement test data in a systematic manner. As a result, the placement test data is very incomplete and is thus not made available to researchers (Email communication with MERI's Data Architect & Manager, Kyle Kwaiser, 2020).

One solution to this problem is to use ELs' end-of-kindergarten ACCESS scores as a proxy control for initial proficiency. A caveat of this solution, however, is that the end-of-kindergarten ACCESS scores are a combined measure of ELs' initial proficiency upon entering school in kindergarten and the gains that they have made during the first year of instruction (kindergarten year) in the program that they are placed into. In other words, the effect of program is still confounded to some extent with the effect of initial proficiency (see Cheung & Slavin, 2012, p. 356 for a discussion). Although this solution is not ideal, it should lead to more a robust

estimate of the effect of instructional programming than not controlling for initial proficiency at all. Another caveat is that the full sample can no longer be used for data analyses due to a multicollinearity problem (i.e., the end-of-kindergarten ACCESS scores are perfectly correlated with that year's proficiency attainment outcomes). To avoid this problem, I selected a subset of the full sample (hereafter, subsample) that only included those students in the three earlier cohorts who had not attained proficiency by the end of the first year of instruction (i.e., kindergarten year). I focused only on the three earlier cohorts because (a) those students had more waves of data available and because (b) they all took ACCESS in kindergarten before WIDA's standard-setting procedure in 2016-2017. As shown in Appendix C, the demographic characteristics of the subsample are very similar those of the full sample (displayed in Table 2 in this Chapter).

It should be noted that although I attempt to control for as much initial variation in EL students as I can, the results should be not interpreted as robust causal estimates. Instead, they should be interpreted as descriptive patterns in time to proficiency by important factors that may affect this time. Results of this study can help generate hypotheses for future research employing designs that can better support causal inferences (see Slavin et al., 2011 for a discussion of how to design an experimental study to examine the effect of school instruction).

#### *7.2.4.1 Research Question 1*

Research question 1 asks about the typical length of time it takes for English learner (EL) children in Michigan public schools to attain proficiency. I used discrete-time survival analysis to answer this research question. Discrete-time survival analysis is specifically designed for questions about time to an event occurrence in a longitudinal context where time is measured discretely in a small number of intervals (like in years in this study; Singer & Willet, 2003). For

this research question, the focal event of interest is *ELP attainment*, an event that marks the time point when an EL attains proficiency as measured by ACCESS.

Historically, survival analysis has been developed and used by medical researchers to measure the duration of time between birth or diagnosis of a disease and the death event. Over the past few decades, the application of survival analysis has been extended beyond biomedical research to other fields, such as criminology (e.g., criminal recidivism), engineering (e.g., mechanical failure after factory rollout), and marketing (e.g., customer retention). Researchers in general education and applied linguistics have adopted this method rather recently. The type of survival analysis used in this study, discrete-time survival analysis, is particularly popular among education researchers given that time is often measured in discrete units such as semesters or academic years (rather than days) in education. Unlike researchers in the medical field, who are often interested in time to a negative event that people want to avoid (e.g., cardiac death), researchers in education and applied linguistics are usually curious about time to a positive event, such as proficiency attainment, reclassification, or graduation (although negative events like high-school dropout are also of interest to education researchers). For example, ELP attainment, the focal event investigated in this study, is a positive event desired by EL educators and researchers. The long tradition of survival analysis in biomedical research has produced a set of technical terminology that has been widely used by researchers within and outside of medical field, such as risk (i.e., eligibility for an event), hazard (i.e., probability of experiencing an event), and survival probability (i.e., cumulative probability of not experience an event). As one may have noticed, all these terms carry a negative connotation. Although such terminology fits the medical context where negative events are of interest, these terms do not necessarily fit analyses of time to socio-educational events, especially in cases where a positive event is under

investigation. For example, we can say someone is at risk for a heart attack, but we do not say some student is at risk for proficiency attainment. In this manuscript, I adopt traditional survival analysis terminology for consistency considerations (all terms borrowed from Singler & Willet, 2003). However, in Chapter 9, I discuss how these terms can be positively rephrased to better serve researchers in social and education sciences. Table 17 displays key terminology that I use to describe survival analysis in this study. In the first column, I show the traditional terminology, and in the second and third columns, I provide adjusted terms that researchers in social science and general education/applied linguistics can adopt in using proficiency attainment as an example.

Table 17. Traditional and adjusted survival analysis terminology for social science, general education, and applied linguistics

Traditional statistical terms	Adjusted terminology for social science	Adjusted terminology for general education/ applied linguistics
Survival analysis	Event history analysis	Proficiency attainment analysis
Risk	Likelihood of experiencing the event	Likelihood of attaining proficiency
Risk set	Set of participants who are eligible to experience the event, but haven't done so yet	Set of learners/students who have not yet attained proficiency
Hazard	Probability of experiencing an event at a given time point or within a given time period	Probability of attaining proficiency at a given time point or within a given time period
Survival probability	Cumulative probability of not experiencing an event at a given time point or within a given time period	Cumulative probability of not attaining proficiency at a given time point or within a given time period
Censored	Participants who left the study without experiencing the event	Learners/students who left the study without attaining proficiency

Survival analysis is superior to other regression-based methods, such as ordinary linear regression, for studying time to proficiency because it accounts for the possibility that some ELs

may never attain English proficiency during the period of data collection. Such students are considered censored because their time-to-event is unknown. In this study, a student may be censored due to two reasons:

- (a) the student had not attained proficiency by the end of the study period (the student either never attains proficiency or attains proficiency after 2018-2019, when data were not available);
- (b) the student had not attained proficiency before they left the study (e.g., the student switched to private schools or home schooling, dropped out, moved out of state, took other ELP assessments).

Instead of being ignored, censored students are incorporated into survival analysis as part of the *risk set*—the group of students who have not attained proficiency but are eligible (i.e., at risk) to do so—for each of the time periods in which they were observed. For the full sample, the complete risk set consists of all students who were classified as ELs when they started kindergarten (time 0). For the subsample, the complete risk set consists of all students who were still classified as ELs by the end of the first year of instruction. The risk set becomes smaller in each subsequent year as ELs become proficient or censored and thus leave the risk set. For example, any student who had not attained proficiency by the end of the 2018-2019 academic year is censored. How the risk set changes over time for the full sample is illustrated in Figure 1.

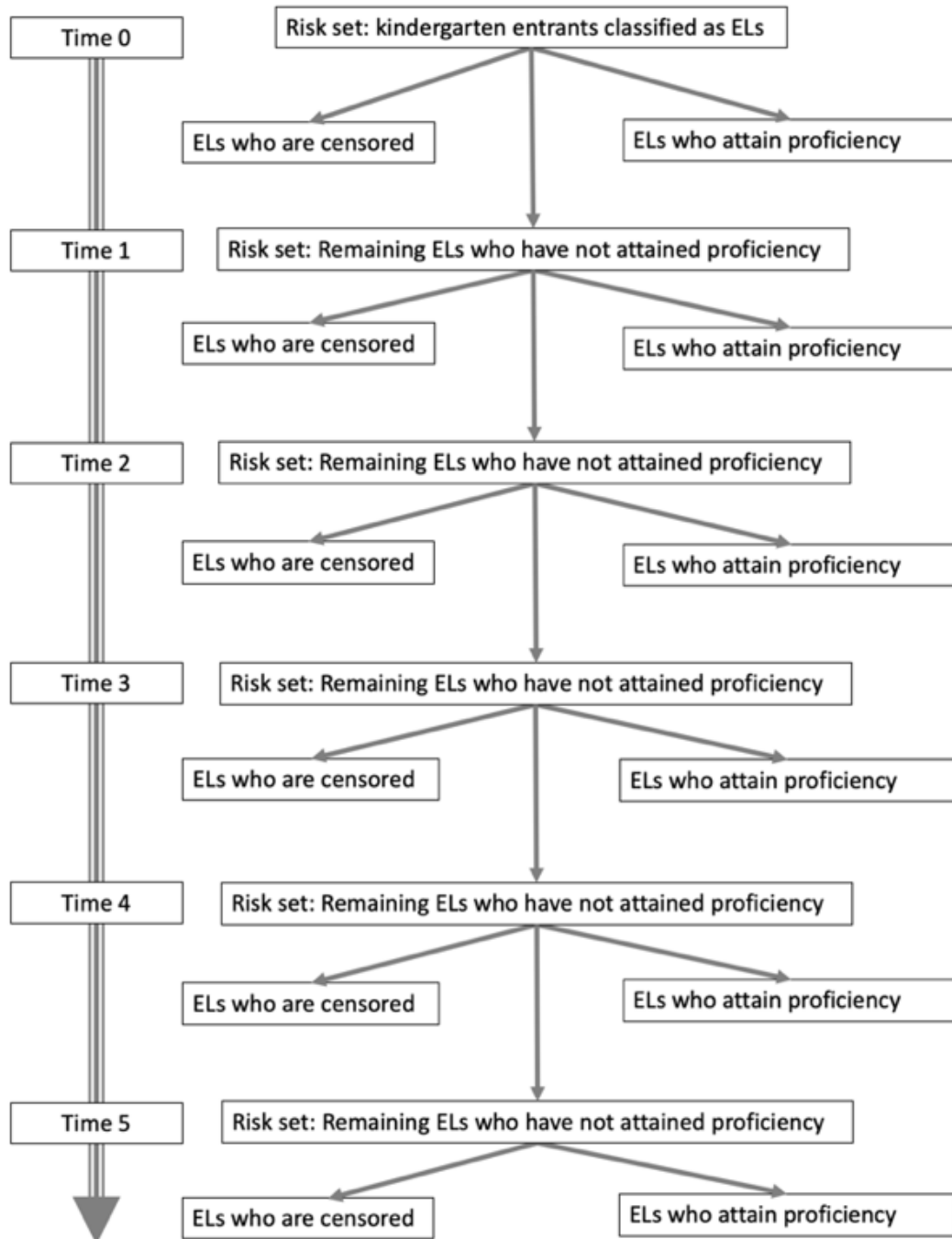


Figure 1. Changes in the risk set from time 0 (year 1 in school) through time 5 (year 6 in school)

Based on the information from the changing risk set, I calculated two functions to answer the first research question: a hazard function and a survival function. The *hazard function*



provides the conditional probability that an EL will attain proficiency during a particular time period provided that the EL has not done so in any earlier time period. This is useful for identifying “risky” time periods when proficiency attainment is especially likely to occur. The *survival function* provides the cumulative probability that a randomly selected EL will *not* attain proficiency after a particular number of years in school (survive = not attain proficiency). This allows one to determine the proportion of ELs who remained in Michigan public schools for a given period of time without attaining proficiency.

To estimate the hazard probability that individual  $i$  attained proficiency during time period  $j$ , I fit the unconditional discrete-time hazard model (in which time was the only covariate):

$$\text{Logit } h(t_{ij}) = \alpha_1 D_{i1} + \alpha_2 D_{i2} + \dots + \alpha_j D_{ij} \quad (\text{Equation 1})$$

where  $D_1$  through  $D_j$  represent a series of dummy-coded indicators for each time period in which the student was observed. The parameter estimates  $\alpha_1$  through  $\alpha_j$  capture the effects of individual time periods on the logit hazard probability of attaining proficiency. The logit function transforms the hazard probability from its original scale bounded between 0 and 1 to the logit scale that covers all the real numbers. This improves the distributional behaviors, prevents inadmissible values, and enables extreme values to be more comparable (Singer & Willet, 2003). The unconditional model is specified as a piecewise function and is used as a baseline to evaluate conditional models with predictors and control variables. This general model specification allows the hazard function to vary across the  $j$  time periods without assuming any specific function forms (e.g., linear or quadratic).

The resulting parameter estimates  $\alpha_1$  through  $\alpha_j$  were first converted to probabilities and were then used to calculate the *survival function*. To calculate the survival probability for time  $j$ ,

I multiplied the survival probability for the previous time period ( $t_{j-1}$ ) by one minus the hazard probability for the current time period ( $t_j$ ):

$$\hat{s}(t_j) = \hat{s}(t_{j-1})[1 - \hat{h}(t_j)] \quad (\text{Equation 2})$$

where  $1 - \hat{h}(t_j)$  represents the probability of surviving time period  $j$  conditional on having survived in the previous time period. For ease of interpretation, I took the complement of the survival probability,  $1 - \hat{s}(t_j)$ , to obtain the cumulative proportion of students who successfully attained proficiency after  $j$  years had elapsed.

Finally, I used the estimated survival probabilities to interpolate the median time to proficiency, which is defined as the time period in which 50% of the student sample were proficient in English. Given that censored cases were allowed to be part of the risk set for those time periods in which they were observed, they contributed to the estimation of the hazard probability in the time periods when they were observed and to the survival probability across all time periods.

#### *7.2.4.2 Research Question 2*

Research question 2 asks about the relationship between the following six factors and ELs' likelihood of proficiency attainment: (a) primary disability status, (b) primary home language, (c) home English use, (d) poverty status, (e) initial instructional program in kindergarten, (f) retention.

I first checked if there was multicollinearity among the six predictors based on their data type. The first type of predictors was nominal, including primary disability status, primary home language, and initial instructional programming kindergarten. The remaining three predictors—home English use, poverty status, and retention—were ordinal, each having two distinct levels (0 and 1). In this study, I assumed that “0” and “1” of the ordinal predictor arose from discretization

(i.e., dividing a spectrum of continuous scores into a finite number of discrete elements) of an underlying continuous variable that was normally distributed. For example, a person's poverty status, or economic status, which is typically defined by the person's (family) income is a normally distributed, continuous variable; however, in this study, I defined poverty dichotomously based on whether the student received subsidized meal. Such a definition discretized an underlying continuous variable into a two-level ordinal variable. In checking multicollinearity, I used crosstabulation to examine the relationship between two normal variables and between a nominal variable and an ordinal variable. I then calculated pairwise polychoric correlation among discretized, ordinal predictors. Polychoric correlation is specifically designed to examine the linear relationship between two discretized, ordinal variables (Ekström, 2011). As I describe in detail in Appendix D, these analyses only indicated noticeable correspondence for two predictors, primary home language and poverty status, and such correspondence was not strong enough to cause any multicollinearity problem in the modeling process.

To answer the second research question, I fit a conditional discrete-time hazard model by introducing the six predictors and the control variables mentioned earlier to the unconditional discrete-time hazard model (Equation 1). The hypothesized conditional model for research question 2 is given as follows:

$$\begin{aligned}
 \text{Logit } h(t_{ij}) = & [\alpha_1 D_{i1} + \alpha_2 D_{i2} + \dots + \alpha_j D_{ij}] \\
 & + \beta_1 X_i + \beta_2 \text{Disability}_i + \beta_3 \text{Language}_i + \beta_4 \text{Program}_i \\
 & + \beta_5 \text{Poverty}_{ij} + \beta_6 \text{HomeEnglish}_{ij} + \beta_7 \text{Retention}_{ij} \\
 & + \beta_8 \text{Language}_i * T_{ij} + \beta_9 \text{Program}_i * T_{ij} + \beta_{10} \text{Poverty}_{ij} * T_{ij} \\
 & + \beta_{11} \text{HomeEnglish}_{ij} * T_{ij} + \beta_{12} \text{Retention}_{ij} * T_{ij} \\
 & + \text{Cohort}_i + \text{SchoolDistrict}_{ij} \\
 & (+ \beta_{13} \text{InitialProficiency}_i + \beta_{14} \text{InitialProficiency}_i * T_{ij})
 \end{aligned}
 \tag{Equation 3}$$

where  $D_1$  through  $D_j$  represent a series of dummy-coded indicators for each time period in which the student was observed,  $X_i$  represents a vector of time-invariant control variables (including *gender*, *age at school entry*, *race/ethnicity*),  $Disability_i$  is an indicator for the student's primary disability type,  $Language_i$  represents the student's primary home language,  $Program_i$  indicates the student's initial program upon school entry,  $Poverty_{ij}$  is a time-varying variable denoting if the student had ever been eligible for free/reduced lunch in time period  $j$ ,  $HomeEnglish_{ij}$  is a time-varying variable representing if the student had ever used English at home in time period  $j$ ,  $Retention_{ij}$  is a time-varying variable denoting if the student had ever been retained in grade in time period  $j$ ,  $T_{ij}$  is a continuous variable representing the number of years that has elapsed since the student entered kindergarten,  $Language_i * T_{ij}$ ,  $Program_i * T_{ij}$ ,  $Poverty_{ij} * T_{ij}$ ,  $HomeEnglish_{ij} * T_{ij}$ ,  $Retention_{ij} * T_{ij}$  are five interaction terms representing the interaction between each of the five predictors and the number of years that has elapsed since the student entered kindergarten,  $Cohort_i$  represents fixed effects for cohort,  $SchoolDistrict_{ij}$  is a time-varying variable representing the fixed effects for school district,  $InitialProficiency_i$  is a time-invariant indicator for ELs initial English proficiency (as measured by their end-of-kindergarten ACCESS scores), and  $InitialProficiency_i * T_{ij}$  represents the interaction between initial proficiency and the number of years that has elapsed since school entry. I put  $InitialProficiency_i$  and its interaction with time in parentheses because these two variables were only included in the model for the subsample. As explained earlier, initial proficiency could not be entered into the model for the full sample because of a multicollinearity problem (i.e., the end-of-kindergarten ACCESS scores are perfectly correlated with that year's proficiency attainment outcomes). Therefore, different conditional models were fitted to the full sample and to the subsample.

#### *7.2.4.3 Research Question 3*

Research question 3 asks about the time it takes ELs to meet five individual proficiency criteria, including a minimum composite score and minimum cut scores on the reading, writing speaking, and listening domains of ACCESS. To answer this research question, I fit five separate unconditional hazard models (as specified in Equation 1) to the data, each to estimate the time it took students to first attain one of the five criteria. I also calculated the survival function and the median time to achieving each individual proficiency criterion based on the hazard estimates.

#### *7.2.4.4 Assumptions Underlying Discrete-time Survival Analysis*

The hazard models postulated above invoke two common assumptions for discrete-time survival analysis. The first assumption is related to censoring. For survival analysis to be valid, censoring must be non-informative (Thompson, 2017). That is, censoring must be “independent of event occurrence and the risk of event occurrence” (Singler & Willet, 2003, p. 318). Recall that students may be censored in this study either because (a) they had not attained proficiency by the end of the study period or because (b) they had not attained proficiency before they left the study. Regardless of the censoring mechanism, I assume that all censoring is non-informative and is unrelated to students’ likelihood of attaining English proficiency. Please see Appendix E *Assumption Testing* for how I tested this assumption.

The second assumption is the proportionality assumption, which stipulates that the effect of a predictor (or a control variable) on the outcome is identical in every time period. In this study, I assume the proportionality assumption is true for all control variables (i.e., gender, race/ethnicity, age, initial proficiency). To take gender as an example, I assume the effect of being female (versus male) remains the same on a student’s likelihood of attaining proficiency (or any other outcomes) in the first year and each subsequent year in which the student was

enrolled in Michigan public schools. I made this assumption for considerations of model parsimony. Meanwhile, I tested if the proportionality assumption is true for each of the following predictors: primary disability type, primary home language, program, poverty status, English use at home, and retention. I first relaxed this assumption for program by including an interaction between program and time ( $Program_i * T_{ij}$ ) in Equation 3. This is because both theory and previous research suggest that compared with students in ESL programs, students in bilingual programs tend to be less likely to attain proficiency in early years when they are receiving greater amounts of instruction in their primary languages but may be more likely to attain proficiency in later years if they benefit from continued exposure to English and transfer from home language and literacy skills (Thompson, 2017; Umansky & Reardon, 2014). The decision to relax the proportionality assumption for program was supported by a statistical test. The test, recommended by Singler and Willet (2003), involves a comparison of the goodness-of-fit between a model that assumed proportionality for program and an otherwise-the-same model that did not make this assumption. Details about assumption testing can be found in Appendix E. I used the same testing procedure to evaluate the appropriateness of the proportionality assumption for each of the remaining predictors. The analyses indicated that the assumption held only for one predictor, primary disability type. I therefore allowed all focal predictors except for primary disability type to have time-varying effects on the outcome by including five interaction terms in Equation 3 (i.e.,  $Language_i * T_{ij}$ ,  $Program_i * T_{ij}$ ,  $Poverty_{ij} * T_{ij}$ ,  $HomeEnglish_{ij} * T_{ij}$ ,  $Retention_{ij} * T_{ij}$ ).

#### 7.2.4.5 Sensitivity Analysis

For research question 2, I conducted additional sensitivity analyses to see if the effects of the predictors of interest were robust to the exclusion of school and cohort controls. Specifically,

three analyses were performed using the conditional discrete-time hazard model as specified in Equation 3, one analysis without school district, one without cohort, and one last analysis without either of the two control variables. These sensitivity analyses help determine if the model used to answer research question 2 was over-specified.

Chapter summary:

- This study investigates three questions related to ELs' time to proficiency and factors affecting this time using six years of data from students who entered Michigan public schools in kindergarten in 2013-2014 through 2018-2019.
- The sample was gender-balanced, consisting of primarily Hispanic or Latino (35.71%), White (36.77%), and Asian (23.28%) racial and ethnic groups. The majority the students were from Spanish- or Arabic-speaking, low-income homes. About 10% of the students were ever in a bilingual program, and a similar percentage of students were ever in a special-education program.
- Discrete-time survival analysis was used to analyze the data, with the focal event being ELs' English proficiency attainment. The unconditional hazard model included discrete time period as the only predictor. The conditional hazard model additionally included six predictors of interest (primary disability type, primary home language, home English use, poverty status, instructional program, retention) and several control covariates.

## CHAPTER 8 RESULTS

### 8.1 Research Question 1

Research question 1 asks about the typical length of time it takes for English learner (EL) children in Michigan public schools to attain proficiency. For this research question, I present the results based on the full sample (i.e., every student who entered school as an EL in kindergarten between 2003-2004 and 2008-2009) because the full sample includes more cohorts and more waves of data (data from 54,146 children total). Results based on the subsample are presented in Appendix F.

The data of the full sample was first examined through a useful tool called the life table. A life table (Table 18) tracks the sampled students' histories of proficiency attainment from the beginning of time (first year in school, or time 0, which in this study was toward the end of the U.S. Kindergarten year) through the end of data collection (time 5, toward the end of the sixth year in school, which in this study was the U.S. 5<sup>th</sup> grade year of elementary schooling, since these children had started school in Kindergarten). The term "life table" obviously comes from the historical application of survival analysis in the health and systems sciences, but is useful in applied linguistics research as well, especially when researchers are investigating second language acquisition within a certain longitudinal slice of the lifespan. Specifically, this table summarizes the number of students present in the risk set at the beginning of each time period (column C), the number of students who attained proficiency during that time period (column D), and the number of students who were censored by the end of that time period (column E). Using these basic counts, the table also displays the sample's hazard of reaching proficiency in each time period (column F), which is equal to the proportion students who attained proficiency during that period.



Table 18. Life table for proficiency attainment of ELs in the full sample, using data from 2013-2014 through 2018-2019

A	B	C	D	E	F
Time	Years since school entry	Beginning total in the risk set	Attained proficiency	Total censored	Hazard
0	1	54146	2062	11580	3.81%
1	2	40504	686	9386	1.69%
2	3	30432	1724	8060	5.67%
3	4	20648	2898	6258	14.04%
4	5	11492	3776	3727	32.86%
5	6	3989	1284	2705	32.19%
<b>Total</b>			<b>12430</b>	<b>41716</b>	

*Note.* A risk set (column C) consists of students who had not attained proficiency by the beginning of a particular time period but were eligible to do so during that time period. Students are censored if they left the study prior to the end of a particular time period without attaining proficiency (column E). Hazard (column F) is equal to the proportion of students who began a given time period in the risk set and subsequently left the risk set because they attained proficiency during that time period.

By design, the hazard data shown in the life table commensurate with the estimates obtained from the unconditional discrete-time hazard model specified in Equation 1 (see Chapter 7). Table 2 displays the model-based hazard estimates (both on the logit and probability scales) and their 95% confidence intervals (CIs). Those estimates were used to calculate the survival probability, or the cumulative proportion of students who have not reached proficiency, and the cumulative proportion of proficient students, which are also presented in Table 19 (last two columns).

Table 19. Results of the unconditional discrete-time hazard model (specified in Equation 1 in Chapter 7) predicting hazard of the full sample attaining proficiency as a function of time, using data from 2013-2014 through 2018-2019

Time	Years since school entry	Logit hazard	SE (for logit hazard)	Hazard	95% CI for hazard		Survival	Cumulative proportion of proficient students
					Upper	Lower		
0	1	-3.2292	0.0225	0.0381	0.0365	0.0397	0.9619	0.0381
1	2	-4.0612	0.0385	0.0169	0.0157	0.0182	0.9456	0.0544
2	3	-2.8125	0.0248	0.0567	0.0541	0.0593	0.8921	0.1079
3	4	-1.8124	0.0200	0.1404	0.1357	0.1452	0.7669	0.2331
4	5	-0.7146	0.0199	0.3286	0.3200	0.3372	0.5149	0.4851
5	6	-0.7451	0.0339	0.3219	0.3076	0.3366	0.3492	0.6508

*Note.* Hazard is equal to the proportion of students who began a given time period in the risk set and subsequently left the risk set because they attained proficiency during that time period. Survival probability is equal to the cumulative proportion of students who had not attained proficiency by the end of a given time period. The proportions in the last two columns account for students who were censored.

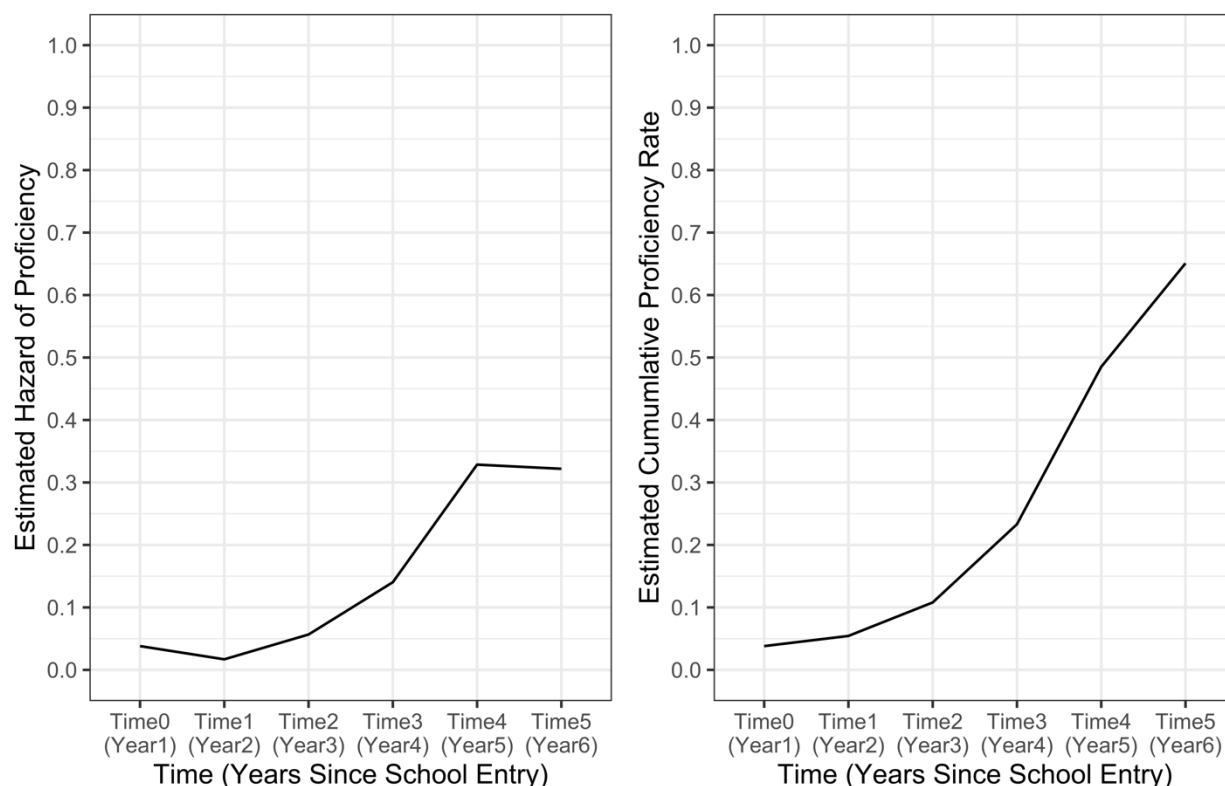


Figure 2. Model-estimated hazard of proficiency (left panel) and cumulative proportion of proficient students (right panel) by time period, using data of the full sample

Examining changes in the estimated hazards (column 5 in Table 19, visualized in the left panel of Figure 2) across time suggests that although the full sample's likelihood of reaching proficiency was initially low when the students were in kindergarten through second grade, the likelihood increased steadily in the upper elementary grades, peaking after the students spent five or six years in Michigan public schools (fourth or fifth grade if the students had not been retained). The observation of a peak in students' likelihood of proficiency attainment toward the end of elementary school is consistent with the findings from some previous studies that tracked ELs' patterns of reclassification through middle and high school years (e.g., Thompson, 2017).

The right panel of the Figure 2 visualizes the cumulative rates of proficiency for the full sample (also listed in the last column of Table 19). Of students who entered kindergarten as ELs, around 50% became proficient in English by the end of their fifth year in Michigan public schools, which equates to the end of grade 4 for a majority of the students. In other words, the median time to English proficiency was roughly five years for students who entered Michigan public schools as ELs in kindergarten. Compared with previous studies that examined ELs' time to reclassification/proficiency, the median time to proficiency found in this study is one to two years longer than similar estimates reported by researchers in New York (Kieffer & Parker, 2016), Massachusetts (Slama, 2014), and Texas (Slama et al., 2017), and is at least one year shorter than those estimates reported by researchers in California (Thompson, 2017; Umansky & Reardon, 2014). The variation in those estimates may be explained by differences in criteria for reclassification/proficiency, English language proficiency tests, student samples, and learning contexts. What is also notable from the cumulative rates of proficiency is that around 35% of students in the full sample were still not proficient in English even after spending six years in

school. These ELs would very likely become long-term ELs (Kieffer & Parker, 2016; Umansky & Reardon, 2014).

## **8.2 Research Question 2**

Research question 2 asks about the relationship between the following six factors and ELs' likelihood of proficiency attainment: (a) primary disability status, (b) primary home language, (c) poverty status, (d) home English use, (e) initial instructional program in kindergarten, (f) retention. First, I present six life tables based on the full sample, with each life table contrasting the event histories of student subgroups defined by a factor of interest. This is to provide an overall picture of each factor's effect on time to proficiency. The life table, however, does not account for the effects of other factors or control variables. To estimate the unique effect of each factor (after controlling for other factors and control variables), I fit two separate conditional discrete-time hazard models (specified in Equation 3 in Chapter 7), one to the full sample and the other to the subsample. Recall that the subsample consists of those students in the three earlier cohorts (2013-2014 kindergarten cohort, 2014-2015 kindergarten cohort, and 2015-2016 kindergarten cohort) who had not attained proficiency by the end of the first year of school (kindergarten). Model results for the subsample are presented in this section following the life tables. I focus on the subsample results because with the subsample, I was able to control ELs' initial English proficiency (as indicated by their end-of-kindergarten ACCESS scores) in estimating the six factors' effects on time to proficiency. Model results for the full sample can be found in Appendix G. Differences between the full and subsample model results are briefly discussed in Appendix G.

### ***8.2.1 Primary Disability Type***

Table 20 displays a life table by primary disability type for the full sample. The life table includes the event histories of four subgroups of students who had different types of primary disabilities. In comparing these subgroups' hazard functions (listed in column 6 of Table 20, visualized in the left panel of Figure 3), one can see that in each time period, the group without disabilities was most likely to become proficient, followed by students with speech or language impairments, and then by students with other disabilities, whereas students with specific learning disabilities were the least likely to become proficient over time. Although the precise locations of peaks and troughs in the hazard function differ slightly across subgroups, their relative temporal positions are similar, suggesting that the effect of primary disability status may not depend on time (see a formal test of this in Appendix E). The cumulative rates of proficiency (listed in column 7 of Table 20, visualized in the right panel of Figure 3) indicate large differences in time to proficiency across the four subgroups. For example, by the end of the sixth year in school, only 9.2% of students with specific learning disabilities became proficient in English, compared to 26.6% of students with other disabilities, 43.5% of students with speech or language impairment, and 70.4% of students without disabilities. These numbers suggest that ELs with specific learning disabilities (90.8%) and those with speech or language impairments (56.5%) were much more likely to become long-term ELs than their peers without disabilities.

Table 20. Life table by primary disability status for the full sample, using data from 2013-2014 through 2018-2019

Primary disability	Time	Years since school entry	Beginning total in the risk set	Attained proficiency	Total censored	Hazard	Cumulative proportion of proficient students
Not disabled	0	1	47734	2006	10284	0.0420	0.0420
	1	2	35444	656	8373	0.0185	0.0598
	2	3	26415	1670	7013	0.0632	0.1190
	3	4	17732	2764	5302	0.1560	0.2560
	4	5	9666	3530	3009	0.3650	0.5280
	5	6	3127	1163	1964	0.3720	0.7040
Specific learning disabilities	0	1	949	0	31	0.0000	0.0000
	1	2	918	0	56	0.0000	0.0000
	2	3	862	1	113	0.0012	0.0012
	3	4	748	4	194	0.0054	0.0065
	4	5	550	13	209	0.0236	0.0300
	5	6	328	21	307	0.0640	0.0921
Speech or language impairment	0	1	3845	43	712	0.0112	0.0112
	1	2	3090	23	665	0.0074	0.0185
	2	3	2402	48	688	0.0200	0.0382
	3	4	1666	117	578	0.0702	0.1060
	4	5	971	197	374	0.2030	0.2870
	5	6	400	83	317	0.2080	0.4350
Other disabilities	0	1	1618	13	553	0.0080	0.0080
	1	2	1052	7	292	0.0067	0.0146
	2	3	753	5	246	0.0066	0.0212
	3	4	502	13	184	0.0259	0.0465
	4	5	305	36	135	0.1180	0.1590
	5	6	134	17	117	0.1270	0.2660

*Note.* A risk set consists of students who had not attained proficiency by the beginning of a particular time period but were eligible to do so in that time period. Students are censored if they left the study prior to the end of a particular time period without attaining proficiency. Hazard is equal to the proportion of students who began a given time period in the risk set and subsequently left the risk set because they attain proficiency during that time period. Statistics in the last two columns do not account for other student, family, or school characteristics.

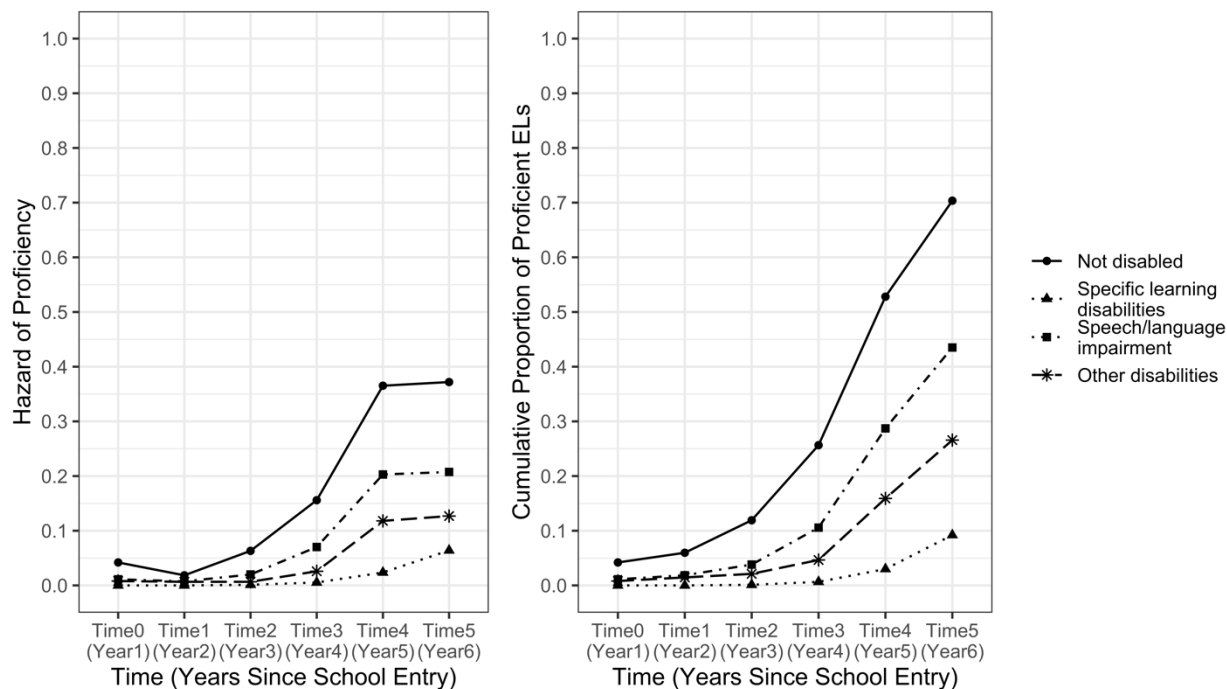


Figure 3. Life-table-derived hazard of proficiency (left panel) and cumulative proportion of proficient students (right panel) by time period and by disability status, using data of the full sample

### 8.2.2 Primary Home Language

A life table by primary home language (Table 21) suggests that EL children from different home language backgrounds varied substantially in their likelihood of attaining proficiency. In general, the hazard functions (listed in column 6 of Table 21) of the 13 home language subgroups displayed three patterns:

- 1) Slow learners: Spanish- and AfroAsiatic-speaking ELs demonstrated the lowest likelihood of becoming proficient in most time periods, with AfroAsiatic ELs showing a slightly higher likelihood of reaching proficiency than Spanish ELs, particularly toward the end of elementary school.
- 2) Fast learners: Chinese, Dravidian, and Japanese/Korean-speaking ELs had the highest likelihood of becoming proficient in most time periods. While the hazard of proficiency increased steadily among Dravidian and Japanese/Korean ELs throughout elementary

school, the hazard function for Chinese learners dropped substantially in the sixth year after a period of steady growth.

- 3) Medium-speed learners: The hazard functions of the remaining home language subgroups fell between those of the slow and fast learners. While the hazard functions of all medium-speed subgroups showed a steady increase during the first five years, only Bengali ELs maintained this trend through the sixth year.

To avoid a crowded presentation, I only visualize the hazard functions for five of the 13 home language subgroups: Spanish, Arabic, Austronian, Bengali, and Japanese/Korean (left panel of Figure 4). These five subgroups were selected for visualization because their hazard functions are representative of the three patterns described above. The fact that the five subgroups' hazard functions crossed with each other in Figure 4 (left panel) suggests that the effect of home language on the likelihood of proficiency attainment may vary across time (see a formal test of this in Appendix E). The right panel of Figure 4 illustrates the cumulative proportion of students attaining proficiency from time 0 through time 5 within each of the five subgroups. As one would expect, slow learners (Spanish ELs) took considerably longer times to attain proficiency than faster learners (Japanese/Korean ELs). For example, only 52.40% of Spanish ELs attained proficiency by the end of the sixth year, compared 93.7% of Japanese/Korean ELs. However, one should note that the hazard functions and cumulative proficiency rates presented in Table 21 could be confounded by other factors that are correlated with both primary home language and the likelihood of proficiency attainment (e.g., poverty status).



Table 21. Life table by primary home language for the full sample, using data from 2013-2014 through 2018-2019

Primary home language	Time	Years since school entry	Beginning total in the risk set	Attained proficiency	Total censored	Hazard	Cumulative proportion of proficient students
Spanish	0	1	19590	250	3911	0.0128	0.0128
	1	2	15429	108	3353	0.0070	0.0197
	2	3	11968	292	3111	0.0244	0.0436
	3	4	8565	632	2784	0.0738	0.1140
	4	5	5149	1301	1831	0.2530	0.3380
	5	6	2017	566	1451	0.2810	0.5240
Afro-Asiatic	0	1	1775	28	346	0.0158	0.0158
	1	2	1401	17	312	0.0121	0.0277
	2	3	1072	39	332	0.0364	0.0631
	3	4	701	64	190	0.0913	0.1490
	4	5	447	121	153	0.2710	0.3790
	5	6	173	66	107	0.3820	0.6160
Albanian	0	1	1061	26	180	0.0245	0.0245
	1	2	855	12	166	0.0140	0.0382
	2	3	677	42	152	0.0620	0.0979
	3	4	483	101	103	0.2090	0.2870
	4	5	279	121	74	0.4340	0.5960
	5	6	84	22	62	0.2620	0.7020
Arabic	0	1	14062	374	2867	0.0266	0.0266
	1	2	10821	127	2414	0.0117	0.0380
	2	3	8280	386	2243	0.0466	0.0829
	3	4	5651	780	1724	0.1380	0.2090
	4	5	3147	1130	1014	0.3590	0.4930
	5	6	1003	328	675	0.3270	0.6590
Austro-nesian	0	1	1363	58	277	0.0426	0.0426
	1	2	1028	45	218	0.0438	0.0845
	2	3	765	79	147	0.1030	0.1790
	3	4	539	120	139	0.2230	0.3620
	4	5	280	157	53	0.5610	0.7200
	5	6	70	35	35	0.5000	0.8600

Table 21 cont'd

Balto-Slavic	0	1	1419	59	249	0.0416	0.0416
	1	2	1111	18	225	0.0162	0.0571
	2	3	868	76	201	0.0876	0.1400
	3	4	591	117	161	0.1980	0.3100
	4	5	313	132	95	0.4220	0.6010
	5	6	86	36	50	0.4190	0.7680
Bengali	0	1	1537	56	321	0.0364	0.0364
	1	2	1160	16	305	0.0138	0.0497
	2	3	839	69	238	0.0822	0.1280
	3	4	532	99	185	0.1860	0.2900
	4	5	248	104	68	0.4190	0.5880
	5	6	76	36	40	0.4740	0.7830
Chinese	0	1	1508	151	466	0.1000	0.1000
	1	2	891	45	229	0.0505	0.1460
	2	3	617	97	133	0.1570	0.2800
	3	4	387	154	76	0.3980	0.5660
	4	5	157	92	29	0.5860	0.8210
	5	6	36	11	25	0.3060	0.8750
Dravidian	0	1	2428	427	595	0.1760	0.1760
	1	2	1406	83	412	0.0590	0.2250
	2	3	911	192	235	0.2110	0.3880
	3	4	484	199	115	0.4110	0.6400
	4	5	170	84	31	0.4940	0.8180
	5	6	55	30	25	0.5450	0.9170
Indo-Iranian	0	1	3068	324	716	0.1060	0.1060
	1	2	2028	70	515	0.0345	0.1360
	2	3	1443	175	359	0.1210	0.2410
	3	4	909	244	250	0.2680	0.4450
	4	5	415	192	113	0.4630	0.7020
	5	6	110	48	62	0.4360	0.8320
Italic/Germanic	0	1	1513	74	343	0.0489	0.0489
	1	2	1096	42	274	0.0383	0.0854
	2	3	780	86	209	0.1100	0.1860
	3	4	485	102	128	0.2100	0.3570
	4	5	255	102	74	0.4000	0.6140
	5	6	79	33	46	0.4180	0.7750

Table 21 cont'd

Japanese/ Korean	0	1	1694	118	577	0.0697	0.0697
	1	2	999	56	375	0.0561	0.1220
	2	3	568	87	206	0.1530	0.2560
	3	4	275	112	71	0.4070	0.5590
	4	5	92	48	24	0.5220	0.7890
	5	6	20	14	6	0.7000	0.9370
Other	0	1	3128	117	732	0.0374	0.0374
	1	2	2279	47	588	0.0206	0.0573
	2	3	1644	104	494	0.0633	0.1170
	3	4	1046	174	332	0.1660	0.2640
	4	5	540	192	168	0.3560	0.5260
	5	6	180	59	121	0.3280	0.6810

*Note.* A risk set consists of students who had not attained proficiency by the beginning of a particular time period but were eligible to do so in that time period. Students are censored if they left the study prior to the end of a particular time period without attaining proficiency. Hazard is equal to the proportion of students who began a given time period in the risk set and subsequently left the risk set because they attain proficiency during that time period. Statistics in the last two columns do not account for other student, family, or school characteristics.

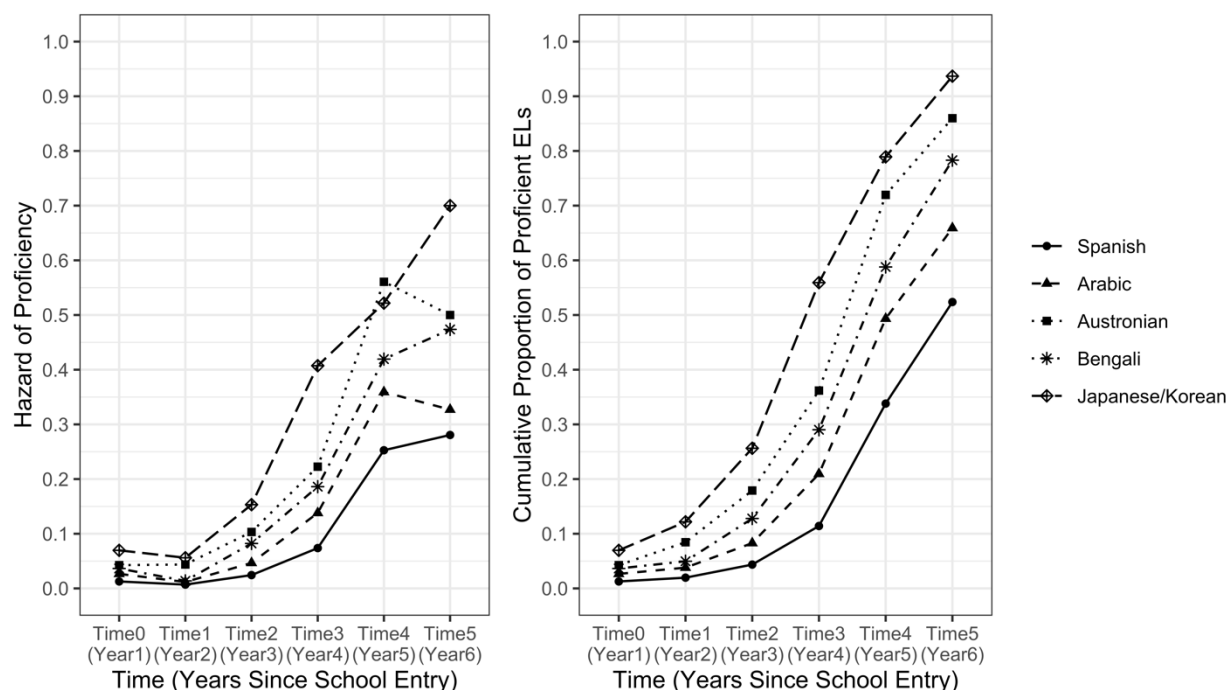


Figure 4. Life-table-derived hazard of proficiency (left panel) and cumulative proportion of proficient students (right panel) by time period and by primary home language, using data of the full sample

### 8.2.3 Poverty Status

In this study, poverty is treated as a time-varying variable indicating whether a student was ever eligible for free/reduced-price lunch between school entry, and a given, later time period. A life table for the full sample by poverty status is presented in Table 22. This table distinguishes between two subgroups of students who differed in their poverty status in each time period: the subgroup (never poor) that was never eligible for free-/reduced-price lunch is shown in the top half of the table, and the subgroup (ever poor) that was ever eligible for free-/reduced-price lunch is included in the bottom half of the table.

Table 22. Life table by poverty status for the full sample, using data from 2013-2014 through 2018-2019

Poverty status	Time	Years since school entry	Beginning total in the risk set	Attained proficiency	Hazard	Cumulative proportion of proficient students
Never poor	0	1	16118	1367	0.0848	0.0848
	1	2	9421	375	0.0398	0.1210
	2	3	6162	828	0.1340	0.2390
	3	4	3543	1125	0.3180	0.4810
	4	5	1411	702	0.4980	0.7390
	5	6	362	184	0.5080	0.8720
Ever poor	0	1	38028	695	0.0183	0.0183
	1	2	31083	311	0.0100	0.0281
	2	3	24270	896	0.0369	0.0640
	3	4	17105	1773	0.1040	0.1610
	4	5	10081	3074	0.3050	0.4170
	5	6	3627	1100	0.3030	0.5940

*Note.* A risk set consists of students who had not attained proficiency by the beginning of a particular time period but were eligible to do so in that time period. Students are censored if they left the study prior to the end of a particular time period without attaining proficiency. Hazard is equal to the proportion of students who began a given time period in the risk set and subsequently left the risk set because they attain proficiency during that time period. Statistics in the last two columns do not account for other student, family, or school characteristics.

The effect of poverty status can be inferred by comparing the hazard function of ever-poor students to that of never-poor students in a given time period. Although the comparison's name

remained the same across time (i.e., never-poor vs. ever-poor students), the students who constituted the comparison in each time period differed. This is because individual students' poverty status may change over time. Specifically, members in the never poor's risk set may switch to become part of the ever poor' risk set in any of the six time periods (note that the method I used to code poverty status does not allow students to switch membership from ever poor to never poor; see a more detailed discussion in Singler & Willet, 2003, p. 430-431).

The two subgroups' hazard functions (visualized in the left panel of Figure 5) suggest a strong poverty effect on students' likelihood of attaining proficiency in each time period. Cumulatively, 87.2% of the never-poor students attained proficiency by the end of the sixth year, compared with 59.4% of the ever-poor subgroup (visualized in the right panel of Figure 5).

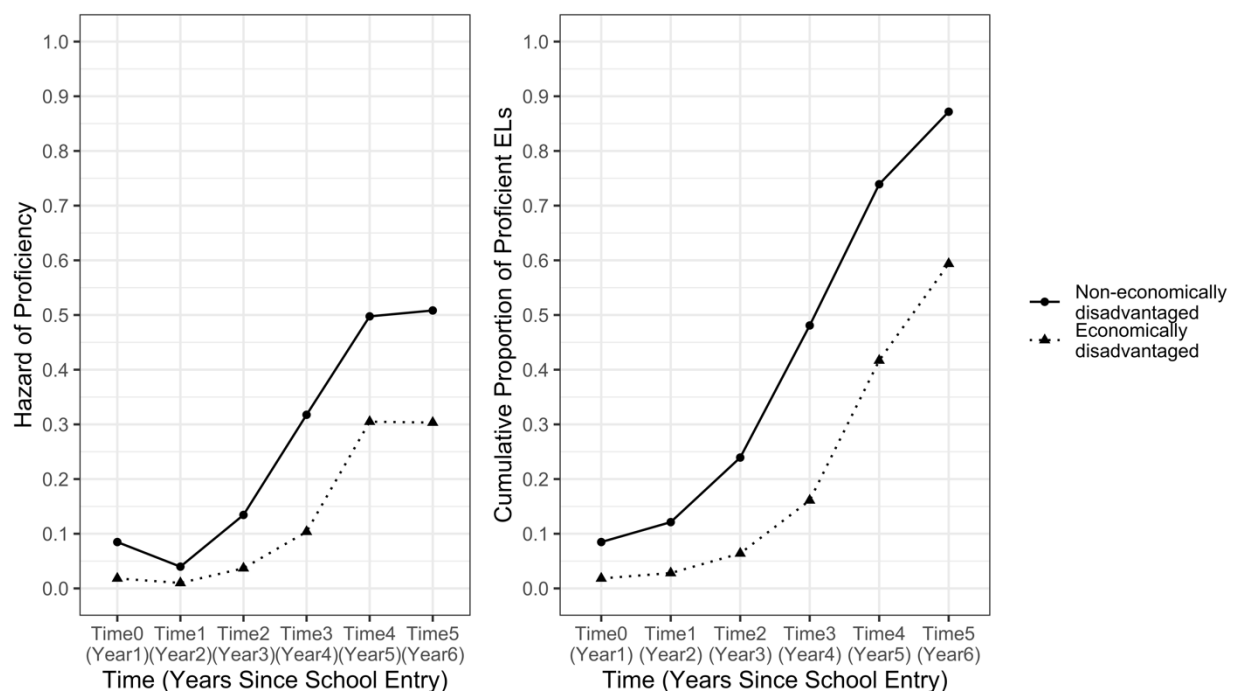


Figure 5. Life-table-derived hazard of proficiency (left panel) and cumulative proportion of proficient students (right panel) by time period and by poverty status, using data of the full sample

### 8.2.4 Home English Use

Home English use is a time-varying variable denoting whether the student ever used English at home between school entry, and a given, later time period. The life table (Table 23) contrasts the event histories of two subgroups of students who had different home-English-use experiences (never use English at home vs. ever use English at home) at each point in time. Because home English use is a time-varying variable, the composition of the two subgroups may change over time when some students switch membership from *never using English at home* to *ever using English at home*. The hazard functions of the two subgroups (visualized in the left panel of Figure 6) suggest that home English use had a small effect on students' likelihood of attaining proficiency, slightly favoring those students who ever used English at home in all time periods except the last one. The cumulative proficiency rates (visualized in the right panel of Figure 6) were similar between the two subgroups over time.

Table 23. Life table by home English use for the full sample, using data from 2013-2014 through 2018-2019

Home English use	Time	Years since school entry	Beginning total in the risk set	Attained proficiency	Hazard	Cumulative proportion of proficient students
Never use English at home	0	1	46998	1680	0.0357	0.0357
	1	2	34117	546	0.0160	0.0512
	2	3	24962	1293	0.0518	0.1000
	3	4	16818	2245	0.1330	0.2200
	4	5	9268	3005	0.3240	0.4730
	5	6	3219	1049	0.3260	0.6450
Ever use English at home	0	1	7148	382	0.0534	0.0534
	1	2	6387	140	0.0219	0.0742
	2	3	5470	431	0.0788	0.1470
	3	4	3830	653	0.1700	0.2930
	4	5	2224	771	0.3470	0.5380
	5	6	770	235	0.3050	0.6790

*Note.* A risk set consists of students who had not attained proficiency by the beginning of a particular time period but were eligible to do so in that time period. Students are censored if they

left the study prior to the end of a particular time period without attaining proficiency. Hazard is equal to the proportion of students who began a given time period in the risk set and subsequently left the risk set because they attain proficiency during that time period. Statistics in the last two columns do not account for other student, family, or school characteristics.

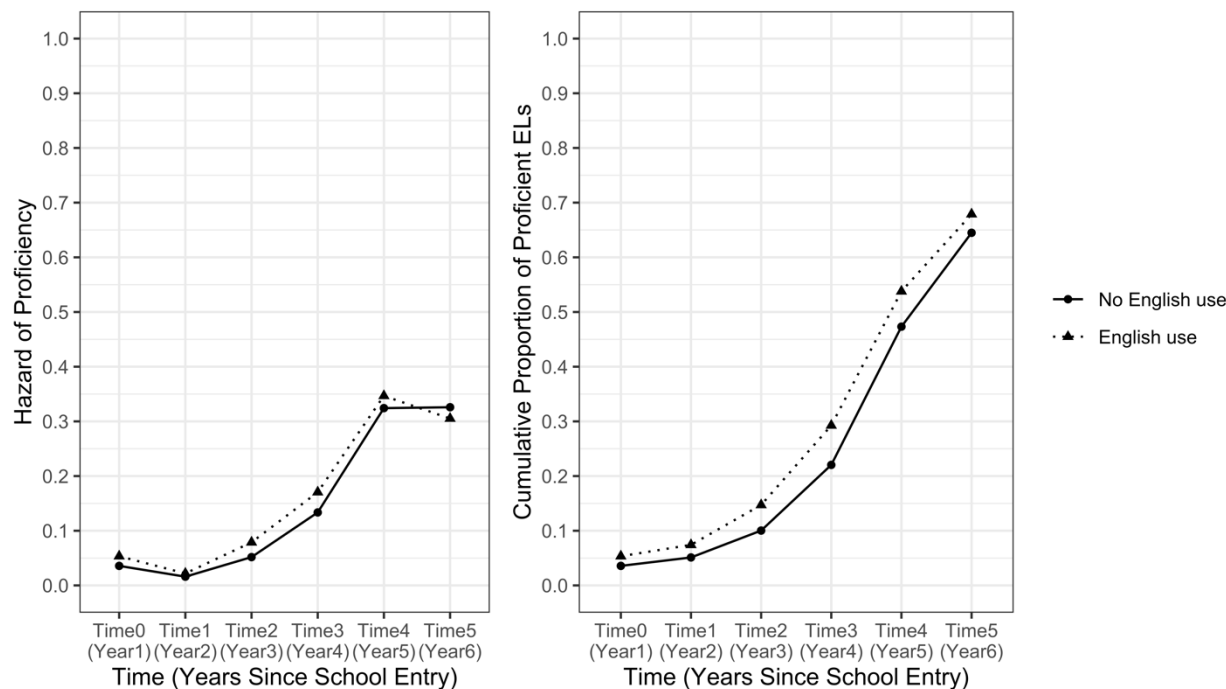


Figure 6. Life-table-derived hazard (likelihood) of proficiency (left panel) and cumulative proportion of proficient students (right panel) by time period and by home English use, using data of the full sample

### 8.2.5 Instructional Program in Kindergarten

Summary statistics in Table 24 suggest that instructional programming also had an impact on ELs' likelihood of proficiency attainment over time. The left panel of Figure 7 illustrates the hazard functions (listed in column 6 in Table 24) of four student subgroups defined by their initial instructional program upon kindergarten entry. The relative levels of the four groups' hazard functions indicate that, in general, students who had no language-learning programs and those placed in an ESL program tended to have a higher likelihood of becoming proficient than those enrolled in a bilingual program in each time period. Similar trends are also seen in the four

groups' cumulative proficiency rates (listed in column 7 of Table 24, visualized in the right panel of Figure 7).

Table 24. Life table by instructional program for the full sample, using data from 2013-2014 through 2018-2019

Program	Time	Years since school entry	Beginning total in the risk set	Attained proficiency	Total censored	Hazard	Cumulative proportion of proficient students
ESL program	0	1	48244	1976	10737	0.0410	0.0410
	1	2	35531	666	8616	0.0187	0.0589
	2	3	26249	1638	7300	0.0624	0.1180
	3	4	17311	2488	5592	0.1440	0.2440
	4	5	9231	3097	3190	0.3350	0.4980
	5	6	2944	963	1981	0.3270	0.6620
Transitional bilingual	0	1	3804	54	423	0.0142	0.0142
	1	2	3327	6	397	0.0018	0.0160
	2	3	2924	52	414	0.0178	0.0335
	3	4	2458	316	380	0.1290	0.1580
	4	5	1762	556	355	0.3160	0.4240
	5	6	851	263	588	0.3090	0.6020
Dual-language bilingual	0	1	1699	9	349	0.0053	0.0053
	1	2	1341	3	337	0.0022	0.0075
	2	3	1001	16	247	0.0160	0.0234
	3	4	738	53	256	0.0718	0.0935
	4	5	429	91	164	0.2120	0.2860
	5	6	174	49	125	0.2820	0.4870
No program	0	1	399	23	71	0.0576	0.0576
	1	2	305	11	36	0.0361	0.0916
	2	3	258	18	99	0.0698	0.1550
	3	4	141	41	30	0.2910	0.4010
	4	5	70	32	18	0.4570	0.6750
	5	6	20	9	11	0.4500	0.8210

*Note.* A risk set consists of students who had not attained proficiency by the beginning of a particular time period but were eligible to do so in that time period. Students are censored if they left the study prior to the end of a particular time period without attaining proficiency. Hazard is equal to the proportion of students who began a given time period in the risk set and subsequently left the risk set because they attain proficiency during that time period. Statistics in the last two columns do not account for other student, family, or school characteristics.



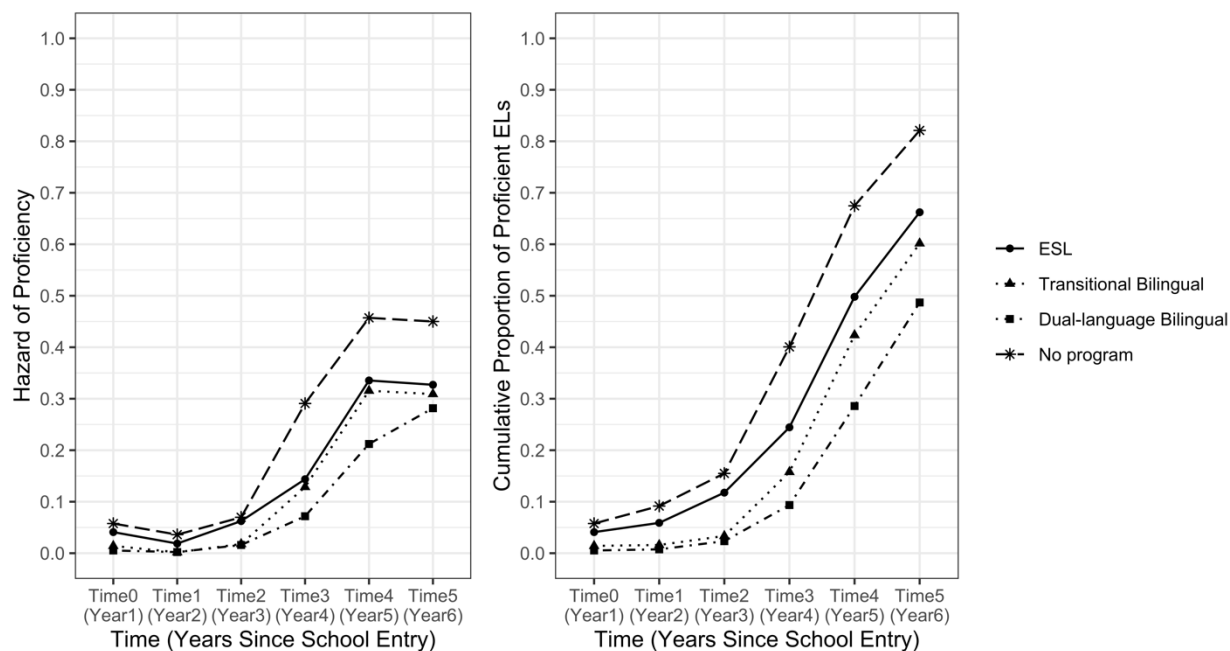


Figure 7. Life-table-derived hazard of proficiency (left panel) and cumulative proportion of proficient students (right panel) by time period and by instructional programming, using data of the full sample

However, one should note that the life table does not account for other factors that may systematically affect EL program placement, such as initial English proficiency, primary home language, and parental preference for ESL programming. Some of these factors are known to correlate significantly with time to proficiency. For example, previous research has found that students with high initial proficiency in English are more likely to be placed in an ESL program than in a bilingual program and are also more likely to become reclassified within a shorter period of time than those with low initial proficiency (Conger, 2010; Thompson, 2017; Umansky & Reardon, 2014). It is, therefore, important to control for initial proficiency and other potential confounding variables when one investigates the effect of instructional program on time to proficiency.

Another notable observation regarding the four groups' hazard functions (left panel of Figure 7) is that although students in dual-language bilingual programs had a lower likelihood of becoming proficient during the first five years, only those students' hazard function maintained

an upward trend throughout the six time periods, whereas the hazards of the other three subgroups all declined substantially entering the sixth year. This observation suggests that the effect of program may have an interaction with time (see a formal test of this assumption in Appendix E).

### 8.2.6 Retention

Retention was operationalized as a time-varying variable denoting whether the student was ever retained between school entry, and a given, later time period. Table 25 presents the event histories for students who were *never retained* and for students who were *ever retained* by time period.

Table 25. Life table by retention for the full sample, using data from 2013-2014 through 2018-2019

Retention	Time	Years since school entry	Beginning total in the risk set	Attained proficiency	Hazard	Cumulative proportion of proficient students
Never retained	0	1	54146	2062	0.0381	0.0381
	1	2	37068	466	0.0126	0.0502
	2	3	27531	1691	0.0614	0.1090
	3	4	18514	2798	0.1510	0.2430
	4	5	10243	3717	0.3630	0.5180
	5	6	3419	1148	0.3360	0.6800
Ever retained	0	1	NA	NA	NA	NA
	1	2	3436	220	0.0640	0.0640
	2	3	2901	33	0.0114	0.0747
	3	4	2134	100	0.0469	0.1180
	4	5	1249	59	0.0472	0.1600
	5	6	570	136	0.2390	0.3600

*Note.* A risk set consists of students who had not attained proficiency by the beginning of a particular time period but were eligible to do so in that time period. Students are censored if they left the study prior to the end of a particular time period without attaining proficiency. Hazard is equal to the proportion of students who began a given time period in the risk set and subsequently left the risk set because they attain proficiency during that time period. Statistics in the last two columns do not account for other student, family, or school characteristics.

Note that a student cannot be retained in their first year of school. This is why a risk set does not exist for the ever-retained subgroup in time 0. Starting in the second year (time 1), members in the never-retained subgroup may switch to the ever-retained subgroup depending on their retention status in a particular year. The two subgroups' hazard functions are visualized in the left panel of Figure 8. One can see that except for in time 1, being ever retained was associated a substantially smaller likelihood of attaining proficiency. The cumulative proportion of proficient students was also lower among students who were ever retained than among students who were never retained (visualized in the right panel of Figure 8). By the end of the sixth year, only 36% of students in the ever-retained group attained proficiency in English, compared with 68% of students in the never-retained group.

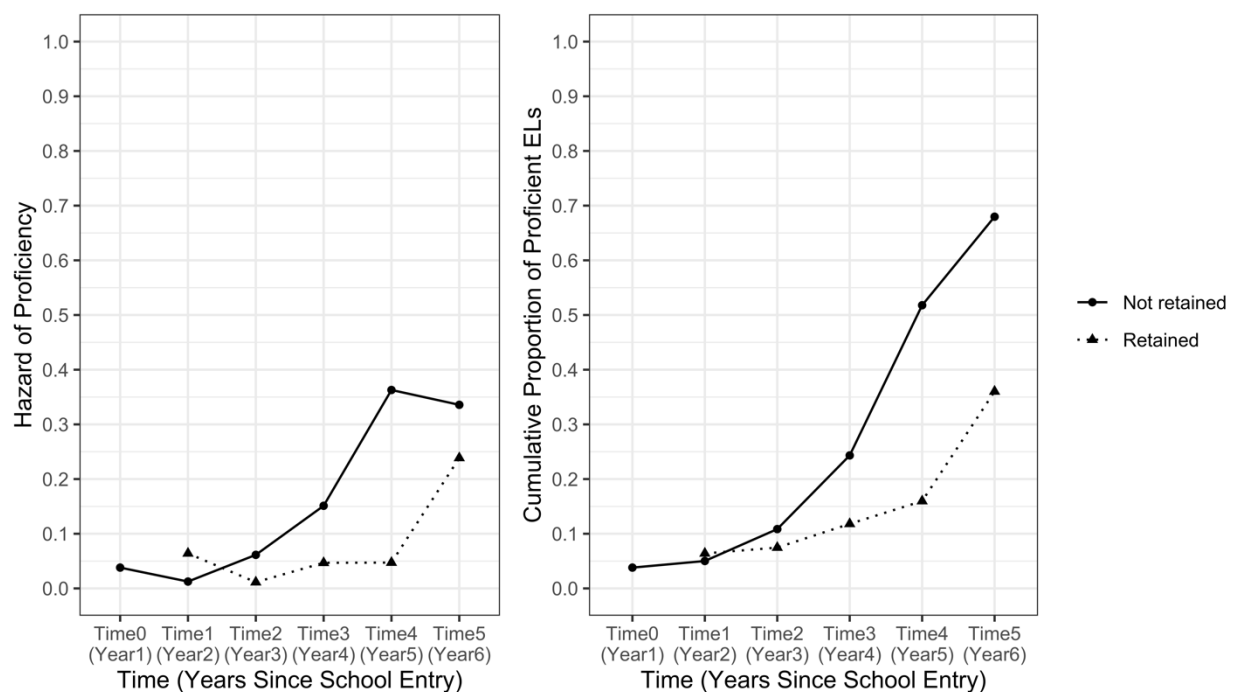


Figure 8. Life-table-derived hazard of proficiency (left panel) and cumulative proportion of proficient students (right panel) by time period and by retention, using data of the full sample

### ***8.2.7 Results for the Conditional Discrete-time Hazard Model (Equation 3 in Chapter 7)***

Table 26 displays the results of fitting the conditional discrete-time hazard model (Equation 3 in Chapter 7) to the data of the subsample. Recall that the subsample consists of those students in the three earlier cohorts (2013-2014 kindergarten cohort, 2014-2015 kindergarten cohort, and 2015-2016 kindergarten cohort) who had not attained proficiency by the end of the first year of school (kindergarten). In other words, students in the subsample all entered the second year of school (grade 1 for most students) without reaching proficiency in English.

For a clearer presentation, only the six predictors and their interactions with time are included in Table 26. (For full results, please see Appendix H). Regression coefficients are reported as logits as well as odds ratios (ORs) to facilitate interpretation. The logit gives the change in the log OR of the outcome for one unit of increase in a particular predictor variable (while holding other variables in the model constant). Positive logit values indicate positive predictor effects on the likelihood of attaining English proficiency. The OR of a predictor describes the change in odds of the outcome for one unit of increase in the predictor (while holding other variables in the model constant). Given that all focal predictors are categorical, one can interpret the OR as the odds that the outcome will occur for a comparison group (predictor value = 1) relative to a reference group (predictor value = 0). An OR of 1 indicates that students in a comparison group have the same odds of attaining proficiency as those in the reference group. ORs greater than 1 indicate that the comparison group is more likely to reach proficiency, whereas ORs less than 1 indicate that the comparison group is less likely to reach proficiency. I focus on ORs in result interpretation because ORs are more intuitive than logits. Based on the

model results, I also extrapolate and report the median time to proficiency for subgroups defined by each predictor.

Table 26. Results for the conditional discrete-time hazard model predicting proficiency hazard for students in the subsample, using data from 2013-2014 through 2018-2019

	Logit	SE (for logit)	z value	p value	OR	95% CI for OR	
						Lower	Upper
Primary disability type (reference = no disabilities)							
Specific learning disabilities	-2.531	0.167	-15.12	<.001***	0.08	0.06	0.11
Speech/language impairments	-0.608	0.058	-10.40	<.001***	0.54	0.49	0.61
Other disabilities	-1.392	0.131	-10.66	<.001***	0.25	0.19	0.32
Primary home language (reference = Spanish)							
AfroAsiatic	-0.261	0.326	-0.80	0.424	0.77	0.41	1.46
Albanian	0.603	0.298	2.02	0.043*	1.83	1.02	3.28
Arabic	0.407	0.178	2.28	0.023*	1.50	1.06	2.13
Austronesian	0.641	0.269	2.38	0.017*	1.90	1.12	3.22
BaltoSlavic	0.316	0.279	1.13	0.257	1.37	0.79	2.37
Bengali	0.786	0.298	2.64	0.008**	2.19	1.22	3.93
Chinese	1.061	0.275	3.85	<.001***	2.89	1.68	4.96
Dravidian	1.407	0.247	5.69	<.001***	4.08	2.52	6.63
IndoIranian	0.612	0.237	2.58	0.010**	1.84	1.16	2.93
Italic/Germanic	0.872	0.265	3.29	<.001***	2.39	1.42	4.02
Japanese/Korean	1.261	0.286	4.42	<.001***	3.53	2.02	6.18
Other home languages	0.552	0.235	2.35	0.019*	1.74	1.10	2.75
Instructional program (reference = ESL)							
Transitional bilingual	-0.339	0.168	-2.02	0.044*	0.71	0.51	0.99
Dual-language bilingual	-0.426	0.356	-1.20	0.231	0.65	0.33	1.31
No program	0.145	0.399	0.36	0.717	1.16	0.53	2.52
Poverty status (reference = never in poverty)							
Retention	-0.508	0.124	-4.09	<.001***	0.60	0.47	0.77
(reference = never retained)							
	2.365	0.185	12.82	<.001***	10.64	7.41	15.28

Table 26 cont'd

Home English use (reference = never used English at home)	0.354	0.108	3.27	0.001**	1.42	1.15	1.76
Interactions							
AfroAsiatic*time	-0.013	0.081	-0.17	0.868	0.99	0.84	1.16
Albanian*time	-0.140	0.080	-1.74	0.081	0.87	0.74	1.02
Arabic*time	-0.146	0.036	-4.01	<.001***	0.86	0.81	0.93
Austronesian*time	-0.049	0.072	-0.67	0.500	0.95	0.83	1.10
BaltoSlavic*time	-0.056	0.076	-0.74	0.461	0.95	0.82	1.10
Bengali*time	-0.130	0.078	-1.66	0.097	0.88	0.75	1.02
Chinese*time	-0.107	0.081	-1.33	0.185	0.90	0.77	1.05
Dravidian*time	-0.398	0.070	-5.70	<.001***	0.67	0.59	0.77
IndoIranian*time	-0.166	0.060	-2.75	0.005**	0.85	0.75	0.95
Italic/Germanic*time	-0.198	0.075	-2.62	0.009**	0.82	0.71	0.95
Japanese/Korean*time	-0.235	0.089	-2.63	0.009**	0.79	0.66	0.94
Other home languages*time	-0.152	0.059	-2.58	0.010**	0.86	0.76	0.96
Transitional bilingual*time	0.072	0.044	1.66	0.096	1.08	0.99	1.17
Dual-language bilingual*time	0.132	0.092	1.43	0.153	1.14	0.95	1.37
No program*time	0.038	0.127	0.30	0.765	1.04	0.81	1.33
Poverty status*time	0.028	0.038	0.73	0.466	1.03	0.95	1.11
Retention*time	-0.789	0.053	-14.93	<.001***	0.45	0.41	0.50
Home English use*time	-0.110	0.031	-3.51	<.001***	0.90	0.84	0.95

Note. p value: <.001\*\*\*, .001-.01 \*\*, .01-.05\*

The model estimated ORs for *primary disability type* reveal that being disabled was associated with a significant and negative effect on ELs' likelihood of attaining proficiency. At each time point, non-disabled students in the subsample were about two times more likely to attain proficiency than their peers with speech or language impairments ( $1/0.54 = 1.85$ ), 12 times more likely to attain proficiency than their peers with specific learning disabilities ( $1/0.08 = 12.50$ ), and four times more likely to attain proficiency than those with other disabilities ( $1/0.25 = 4.00$ ). Figure 9 plots the estimated cumulative proficiency rates by primary disability status (please see Appendix I for more on the graphing method). By the end of the sixth year (or grade

5), 70.4% of non-disabled students in the subsample were expected to reach proficiency, compared with 57.5% of students with speech or language impairments, 19.8% of students with specific learning disabilities, and 40.3% of students with other disabilities. The median time to proficiency for non-disabled students in the subsample was 4 years, which was about half a year sooner than the median time for their peers with speech or language impairments. The median time to proficiency could not be extrapolated for students with specific learning disabilities because it was predicted to be longer than 6 years, which is the maximum number of years for which students were observed in this study.

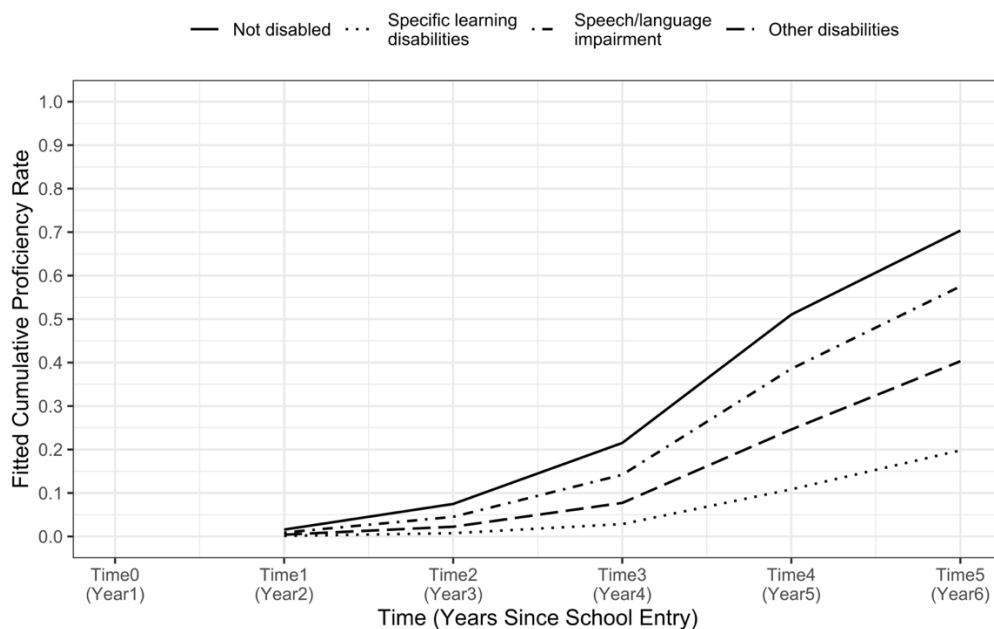


Figure 9. Fitted cumulative proficiency rate by time period and primary disability status, using data of the subsample

Because *primary home language* was assumed to have a time-varying effect on students' likelihood of attaining proficiency, the main effects of primary-home-language variables should be interpreted in combination with their interactions with time. Among the 13 home language groups, Spanish was selected as the reference group and was compared with each of remaining language groups with regard to their likelihood of becoming proficient at each time point. The

main-effect coefficients for primary home language indicate that all non-reference groups except for AfroAsiatic were more likely to attain proficiency than Spanish ELs in early elementary years. However, the interaction terms indicate that those non-reference groups' early "advantages" over the Spanish group became increasingly smaller in later elementary years, with some groups even became less likely to attain proficiency than Spanish. For example, Arabic ELs were 1.50 times more likely to attain proficiency than Spanish ELs in the first year of school (time 0), but they gradually lost their initial advantage (as evidenced by the negative interaction between Arabic and time) and were eventually surpassed by Spanish ELs. By the end of the sixth year (time 5), Spanish ELs were 1.38 ( $1/\exp[(0.407-5*0.146)]$ ) times more likely to attain proficiency than their Arabic peers.

The results for primary home language are most easily interpreted visually. Figure 10 and Figure 11 illustrate the estimated cumulative proficiency rates by home language group (for clarity considerations, the top five home language groups and "other" are included in Figure 10, and the top seven home language family groups are included in Figure 11). Overall, Chinese ELs were estimated to have the highest cumulative probability of reaching proficiency at all time points, whereas AfroAsiatic ELs were estimated to have the lowest cumulative probability of doing so. Specifically, by the end of the sixth year,

- 68.2% of Spanish ELs were expected to attain proficiency;
- 62.0% AfroAsiatic ELs were expected to attain proficiency;
- 68.2% Albanian ELs were expected to attain proficiency;
- 63.9% Arabic ELs were expected to attain proficiency;
- 75.8% Austronesian ELs were expected to attain proficiency;
- 69.6% BaltoSlavic ELs were expected to attain proficiency;



- 72.3% Bengali ELs were expected to attain proficiency;
- 78.4% Chinese ELs were expected to attain proficiency;
- 62.6% Dravidian ELs were expected to attain proficiency;
- 66.2% IndoIranian ELs were expected to attain proficiency;
- 68.7% Italic/Germanic ELs were expected to attain proficiency;
- 72.9% Japanese/Korean ELs were expected to attain proficiency;
- 66.2% other ELs were expected to attain proficiency.

In terms of median time to proficiency, Chinese ELs were expected to have the shortest median time of four and a half years, whereas AfroAsiatic ELs were expected to have the longest median time of five and a half years. Other groups' median times to proficiency fell somewhere in between these two estimates.

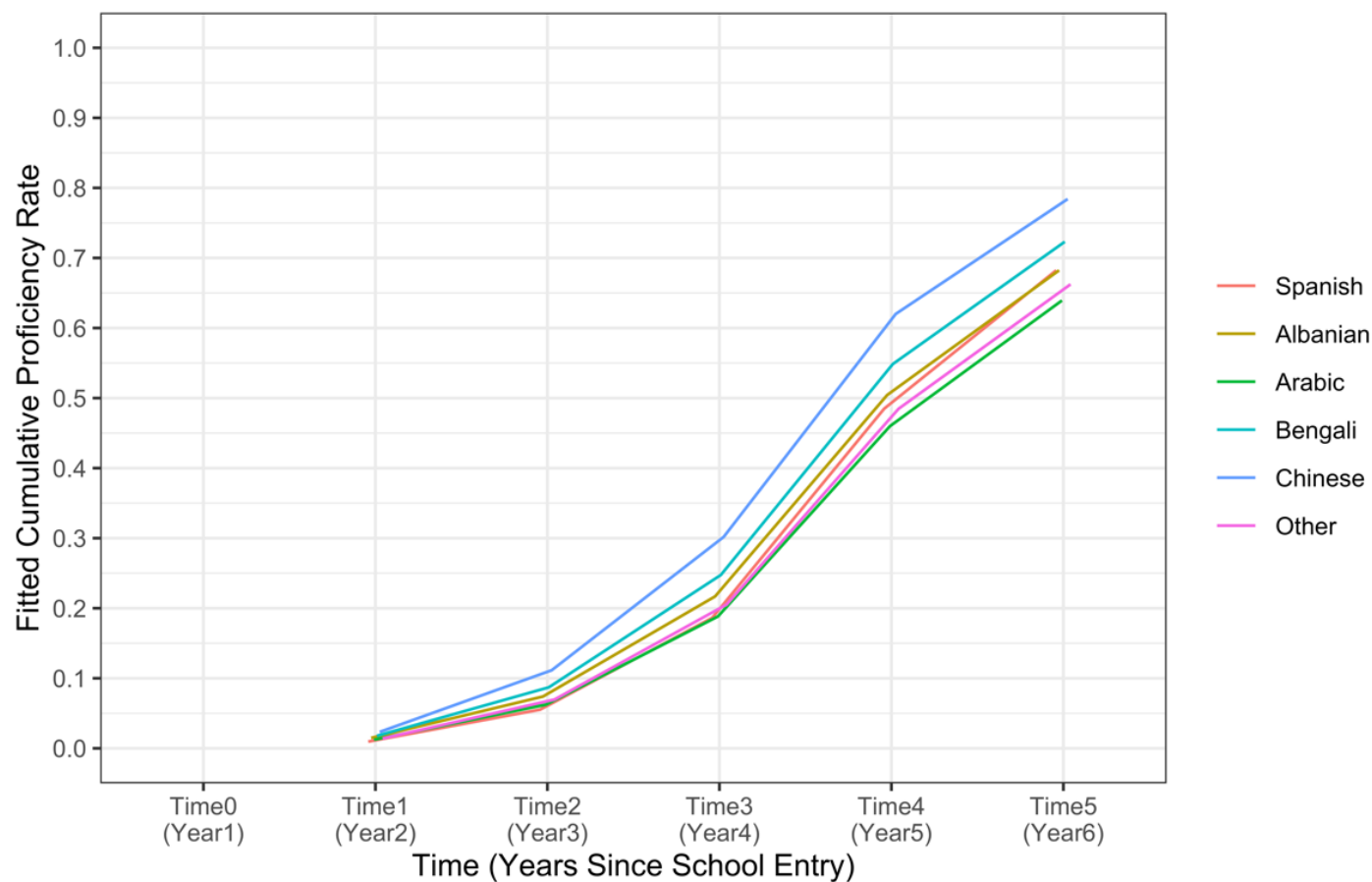


Figure 10. Fitted cumulative proficiency rate by time period and primary home language (including Albanian, Arabic, Bengali, Chinese, Spanish, and Other), using data of the subsample

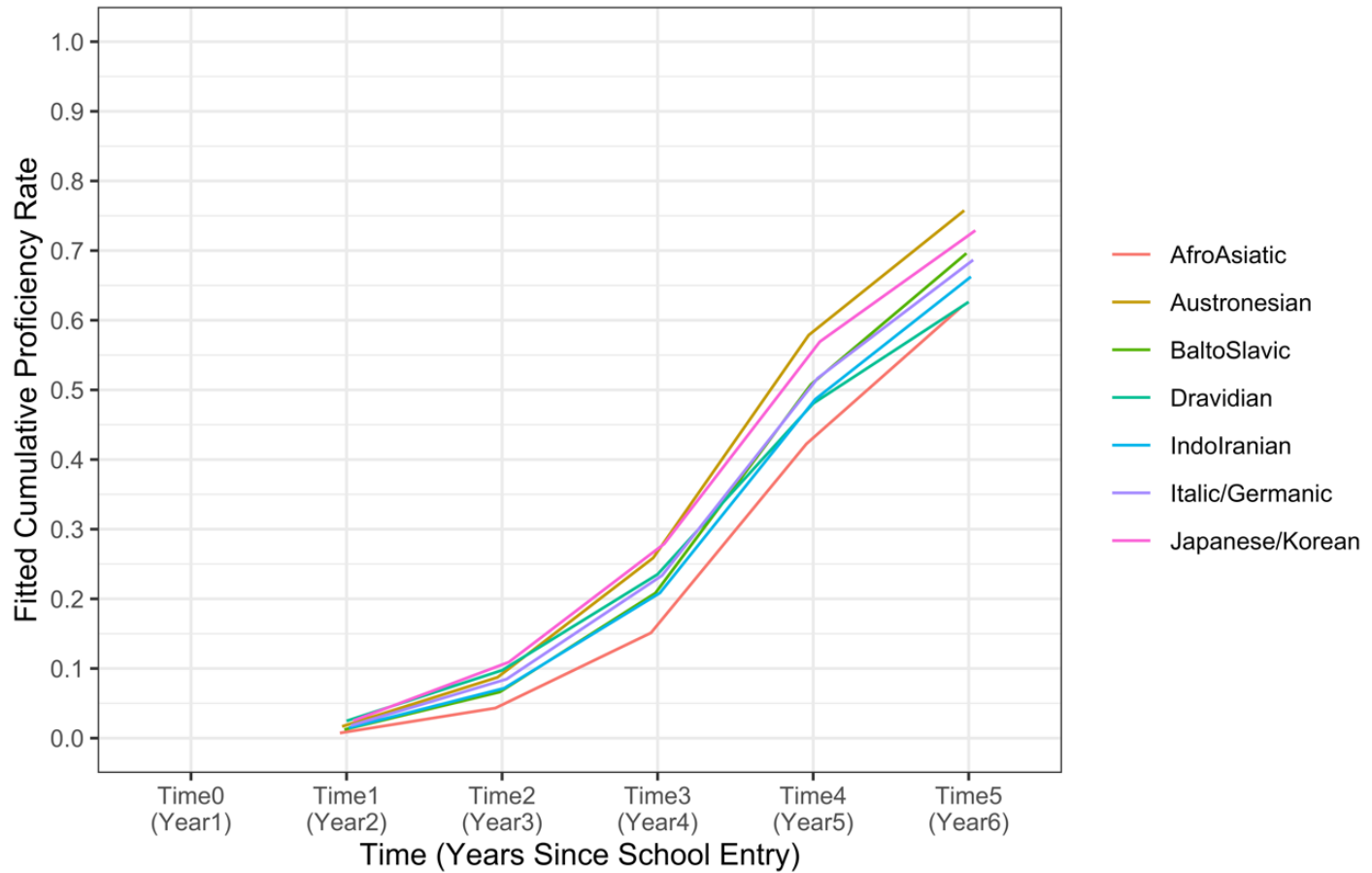


Figure 11. Fitted cumulative proficiency rate by time period and primary home language (including AfroAsiatc, Austronesian, BaltoSlavic, Dravidian, IndoIranian, Italic/Germanic, Japanese/Korean), using data of the subsample

*Poverty status* had a negative, statistically significant main effect and a non-significant interaction with time. This indicates that the state of ever being in poverty was associated with a constant, negative effect on the likelihood of attaining proficiency over time. At each time point, the odds that a student who was never in poverty attained proficiency was nearly two times ( $1/0.6 = 1.68$ ) the odds that a peer who was ever in poverty would do the same. In terms of median time to proficiency, 50% of those students who were never in poverty attained proficiency in slightly more than four and a half years (visualized in Figure 12). The median time to proficiency was around half a year longer for students who were ever in poverty. By the end

of the sixth year, 73.3% of students who were never in poverty reached proficiency, compared to 66.1% of students who were ever in poverty.

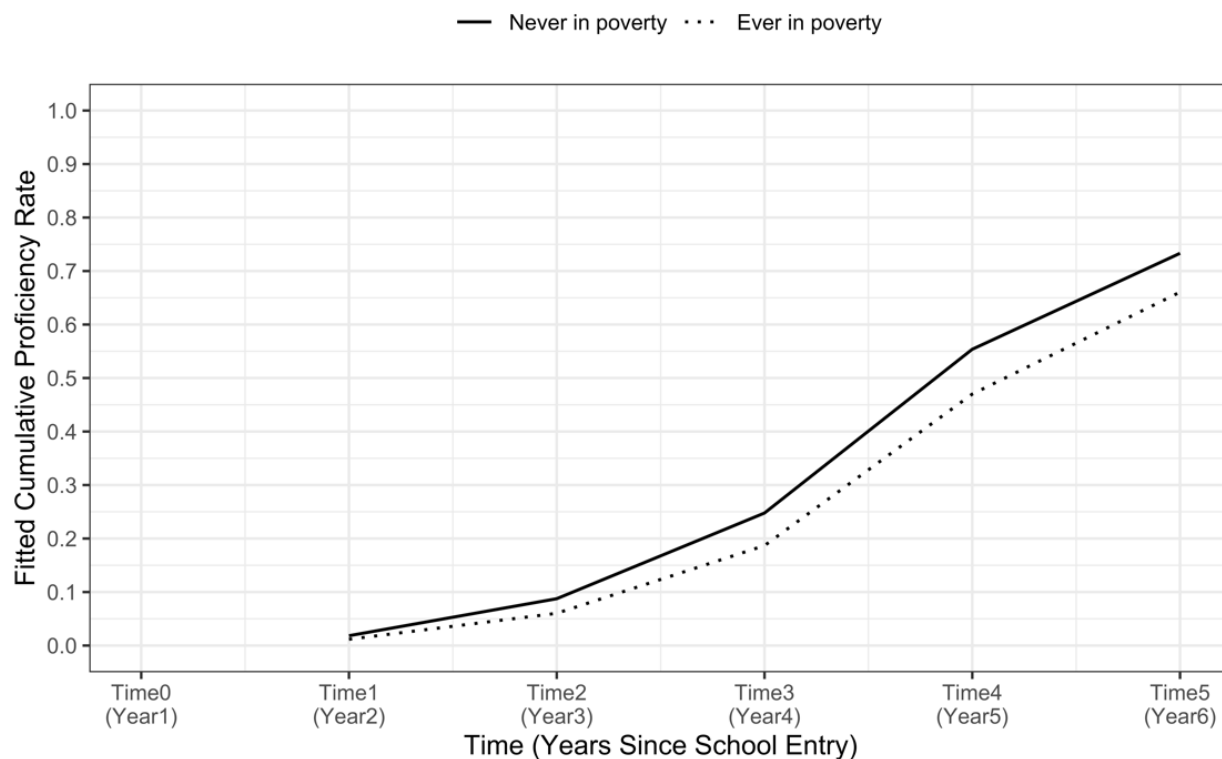


Figure 12. Fitted cumulative proficiency rate by time period and poverty status, using data of the subsample

*Home English use*, operationalized as whether the student ever used English at home, had a small positive (but statistically significant) main effect on the likelihood of attaining proficiency. The effect of this variable was strongest in the early elementary school years and gradually declined over time (visualized in Figure 13), as evidenced by the negative interaction between home English use and time. When students first entered school in kindergarten (time 0), those who ever used English at home were 1.42 times more likely to attain proficiency than their peers who never used English at home. This trend shift over time in favor of students who never used English at home. In the sixth year (time 5), students who never used English at home were 1.22 times ( $1/\exp[(0.354-5*0.110)]$ ) more likely to reach proficiency than their peers who ever

used English at home. Nonetheless, the differences between the two groups' cumulative proficiency rates were small in all time periods (Figure 13); for example, the cumulative proficiency rate was around two percentage points higher in students who never used English at home (67.6%) than in those who ever used English at home (65.3%) in the sixth year (time 5).

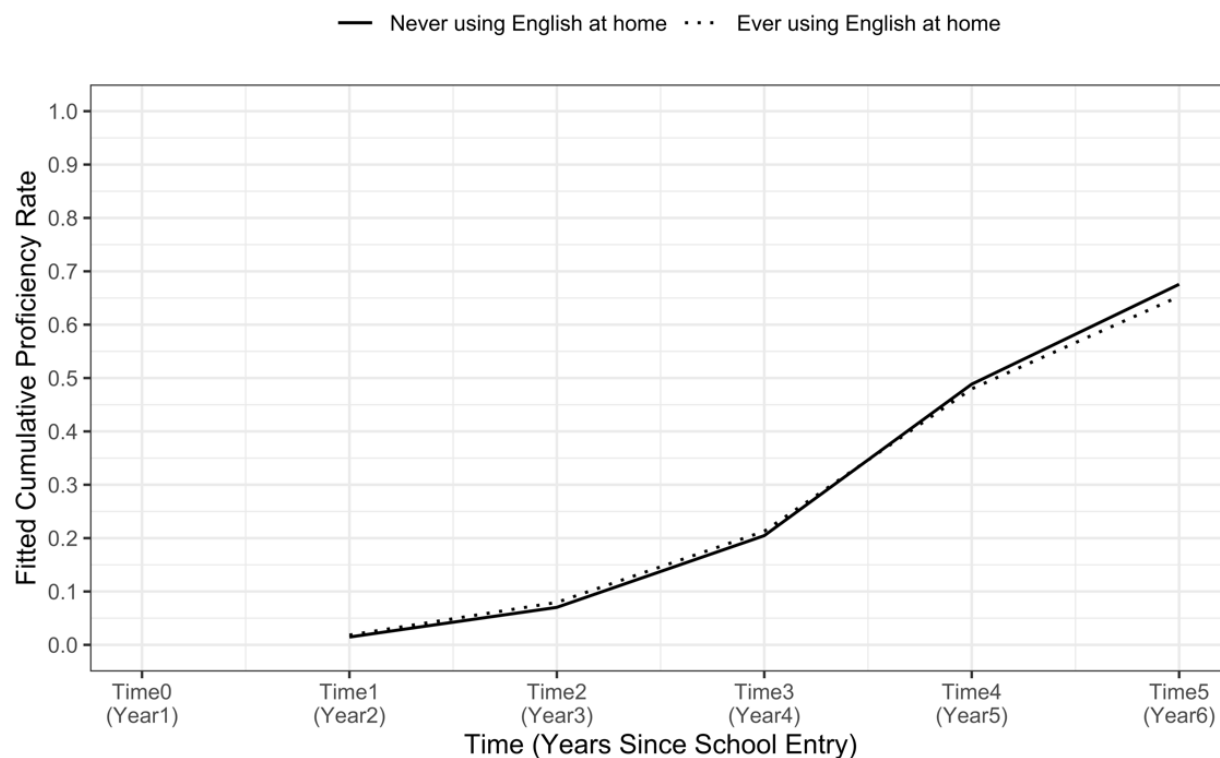


Figure 13. Fitted cumulative proficiency rate by time period and home English use, using data of the subsample

*Instructional programming* was coded as a time-invariant variable but was allowed to have a time-varying effect on the likelihood of attaining proficiency. ESL program was selected as the reference category and was compared with transitional bilingual program, dual-language bilingual program, and no program. After controlling for other predictors and control variables, instructional programming had a negligibly small effect on the hazard of proficiency attainment in each time period. The only significant predictor was the main effect for transitional bilingual program, with its interaction with time approaching statistical significance. Neither the main

effect nor the interaction term was significant for dual-language bilingual program or no program. Figure 14 visualizes the model results, reporting the estimated cumulative proportion of the subsample attaining proficiency by instructional programming. As one can see, students in the four groups demonstrated similar patterns of time to proficiency over time. Figure 14 also shows a couple of interesting but *insignificant* trends. In general, students with no program were slightly more likely to attain proficiency than those receiving specialized language instruction. Regarding students participating in the three instructional programs, those in ESL programs were slightly more likely to attain proficiency than those in bilingual programs in early elementary years, but students in bilingual programs caught up and closed the gaps over time. In particular, the cumulative proficiency rate of dual-language bilingual program (69.8%) surpassed that of ESL program (67.1%) by a small margin of 3 percentage points by the end of the elementary school.

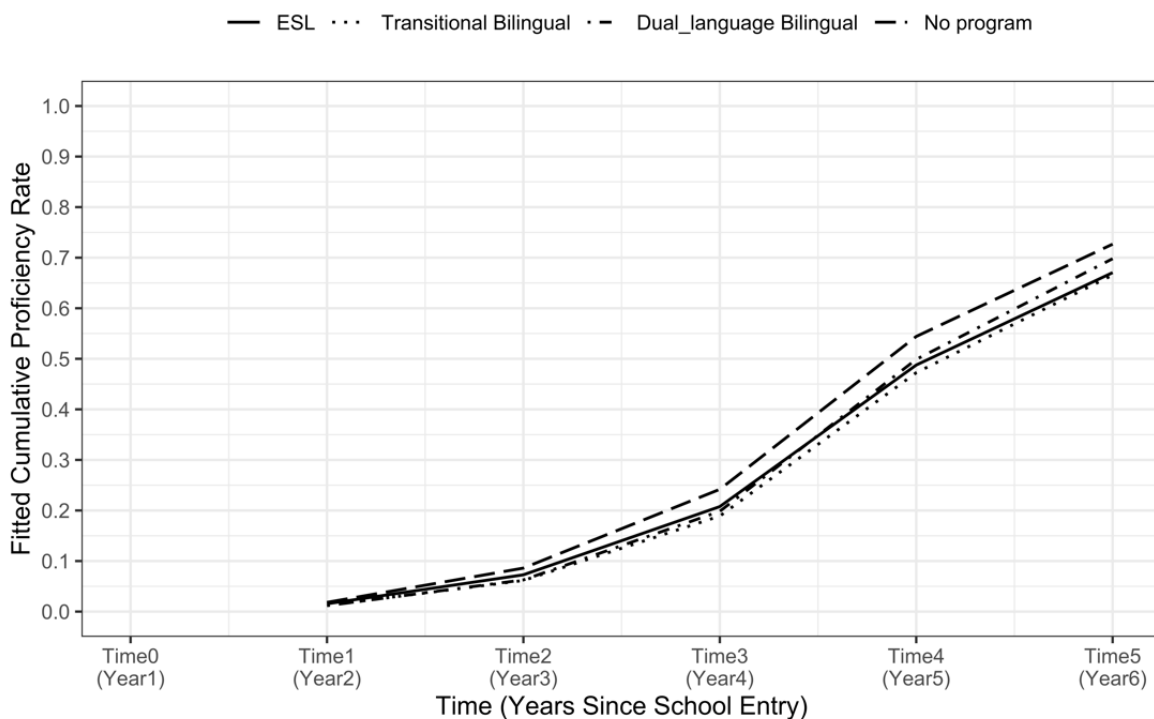


Figure 14. Fitted cumulative proficiency rate by time period and instructional programming, using data of the subsample

*Retention* (i.e., whether the student was ever retained) significantly predicted students' likelihood of attaining proficiency with a positive main effect and a negative interaction with time. Both the strength and direction of the effect of ever being retained on the likelihood of proficiency changed over time (visualized in Figure 15). Specifically, retention had a moderate and positive effect in the second year of school (i.e., the first time when students could be retained), with students who were ever retained being 4.84 times ( $\exp[2.365-0.789]$ ) more likely to attain proficiency than those who were never retained. The effect of retention declined rapidly over time and eventually became moderate and negative in the sixth year (time 5) in which students who were never retained were about 4.85 times ( $1/\exp[2.365-5*0.789]$ ) more likely to attain proficiency than those who were ever retained. Figure 15 illustrates the estimated cumulative proficiency rates by students' retention status. Although a larger proportion of students who were ever retained reached proficiency in early elementary years, starting in the fourth year, students who were never retained demonstrated a higher cumulative proficiency rate than their peers who were ever retained. In the sixth year, the cumulative proficiency rate was almost 20 percentage points higher in students who were never retained (69.3%) than in those who were ever retained (50.3%). Regarding the median time to proficiency, students who were never retained were expected to reach proficiency around 1 year sooner than their peers who were ever retained.

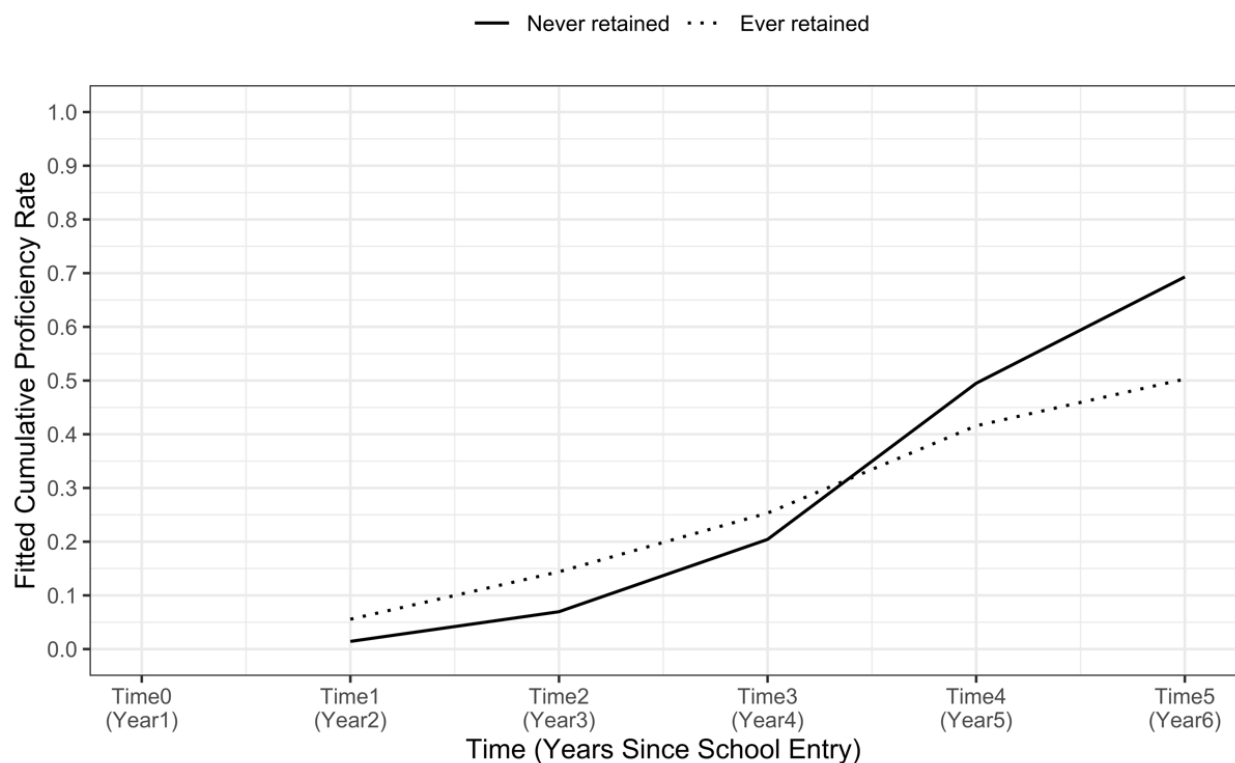


Figure 15. Fitted cumulative proficiency rate by time period and retention, using data of the subsample

Sensibility analyses indicated that the effects of the six predictors of interest were robust to the inclusion or exclusion of school district and cohort controls. Results for sensitivity analyses are reported in Appendix J.

### 8.3 Research Question 3

As described in Chapter 6, ELs in Michigan must meet a variety of proficiency criteria to be considered proficient on ACCESS. The criteria used by the Michigan Department of Education (MDE) for 2013-2014 through 2018-2019 included: (a) reaching a designated minimum composite score on ACCESS (2013-2019); (b) reaching designated cut scores on the reading and writing domains of ACCESS (2013-2019); (c) reaching designated cut scores on the speaking and listening domains of ACCESS (2013-2016).



Research question 3 asks about the time it takes ELs to meet individual proficiency criteria. I used the full sample (i.e., every student who entered school as an EL in kindergarten between 2003-2004 and 2008-2009) to answer this question because the full sample includes more cohorts and more waves of data. Results for the subsample can be found in Appendix K. A life table that tracks how likely ELs were to reach each of the five criteria (i.e., composite score, reading, writing, listening, and speaking) is presented in Table 27.

Table 27. Life table for attaining individual proficiency criteria for ELs in the full sample, using data from 2013-2014 through 2018-2019

Criterion	Time	Years since school entry	Beginning total in the risk set	Reached criterion	Total censored	Hazard
Listening	0	1	28044	16809	4311	0.5994
	1	2	6924	2921	1780	0.4219
	2	3	2223	1579	644	0.7103
Speaking	0	1	28044	9819	6961	0.3501
	1	2	11264	4198	3563	0.3727
	2	3	3503	1710	1793	0.4882
Reading	0	1	54146	12206	9251	0.2254
	1	2	32689	12816	5054	0.3921
	2	3	14819	5215	3038	0.3519
	3	4	6566	1540	2032	0.2345
	4	5	2994	991	1100	0.3310
	5	6	903	218	685	0.2414
Writing	0	1	54146	3018	11327	0.0557
	1	2	39801	754	9204	0.0189
	2	3	29843	2679	7554	0.0898
	3	4	19610	4757	5181	0.2426
	4	5	9672	4084	2573	0.4222
	5	6	3015	1434	1581	0.4756
Composite	0	1	54146	5478	10609	0.1012
	1	2	38059	3646	8268	0.0958
	2	3	26145	3397	6905	0.1299
	3	4	15843	1516	5363	0.0957
	4	5	8964	2836	3091	0.3164
	5	6	3037	885	2152	0.2914

*Note.* A risk set consists of students who had not attained proficiency by the beginning of a particular time period but were eligible to do so during that time period. Students are censored if they left the study prior to the end of a particular time period without attaining proficiency. Hazard is equal to the proportion of students who began a given time period in the risk set and subsequently left the risk set because they attained proficiency during that time period.

Discrete-time survival analysis (as specified in Equation 1 in Chapter 7) was also performed to calculate hazard and survival functions for each proficiency criterion (see Table 28). By design, the hazard statistics shown in the life table are the same as the model estimates displayed in Table 28. For the ease of interpretation, the cumulative proportion of students reaching each of the five criteria is visualized in Figure 16.

Table 28. Results of the discrete-time hazard model (specified in Equation 1 in Chapter 7) predicting hazard of the full sample reaching individual proficiency criteria as a function of time, using data from 2013-2014 through 2018-2019

Criterion	Time	Years since school entry	Hazard	95% CI for hazard		Survival	Cumulative proportion reaching criterion
				Lower	Upper		
Listening	0	1	0.5994	0.5936	0.6051	0.4006	0.5994
	1	2	0.4219	0.4103	0.4335	0.2316	0.7684
	2	3	0.7103	0.6911	0.7288	0.0671	0.9329
Speaking	0	1	0.3501	0.3446	0.3557	0.6499	0.3501
	1	2	0.3727	0.3638	0.3817	0.4077	0.5923
	2	3	0.4882	0.4716	0.5047	0.2087	0.7913
Reading	0	1	0.2254	0.2219	0.2290	0.7746	0.2254
	1	2	0.3921	0.3868	0.3974	0.4709	0.5291
	2	3	0.3519	0.3443	0.3596	0.3052	0.6948
	3	4	0.2345	0.2244	0.2449	0.2336	0.7664
	4	5	0.3310	0.3144	0.3481	0.1563	0.8437
Writing	5	6	0.2414	0.2146	0.2704	0.1186	0.8814
	0	1	0.0557	0.0538	0.0577	0.9443	0.0557
	1	2	0.0189	0.0177	0.0203	0.9264	0.0736
	2	3	0.0898	0.0866	0.0931	0.8432	0.1568
	3	4	0.2426	0.2366	0.2486	0.6387	0.3613
	4	5	0.4222	0.4124	0.4321	0.3690	0.6310
Composite	5	6	0.4756	0.4578	0.4935	0.1935	0.8065
	0	1	0.1012	0.0987	0.1037	0.8988	0.1012

Table 28 cont'd

1	2	0.0958	0.0929	0.0988	0.8127	0.1873
2	3	0.1299	0.1259	0.1341	0.7071	0.2929
3	4	0.0957	0.0912	0.1004	0.6395	0.3605
4	5	0.3164	0.3068	0.3261	0.4372	0.5628
5	6	0.2914	0.2755	0.3078	0.3098	0.6902

*Note.* Hazard is equal to the proportion of students who began a given time period in the risk set and subsequently left the risk set because they attained proficiency during that time period. Survival probability is equal to the cumulative proportion of students who had not attained proficiency by the end of a given time period. The proportions in the last two columns account for students who were censored.

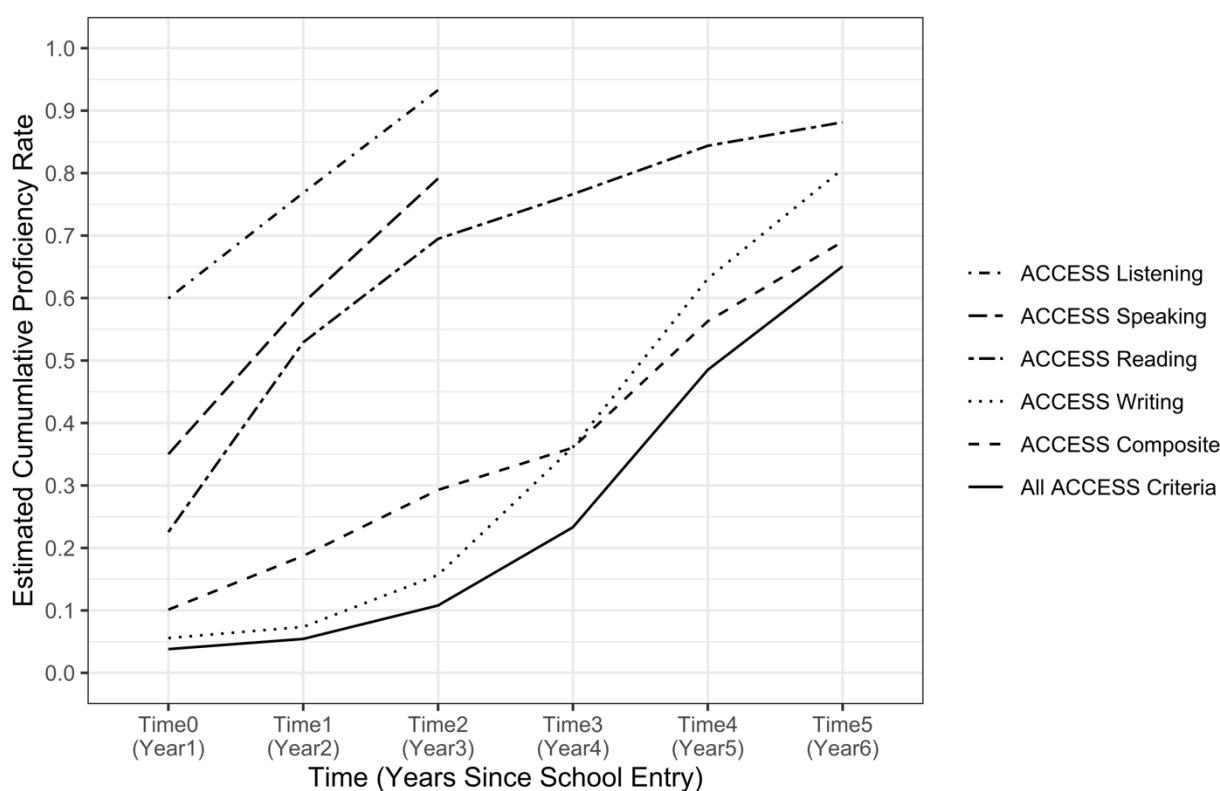


Figure 16. Fitted cumulative proficiency rates for each of the five proficiency criteria (i.e., ACCESS listening, ACCESS speaking, ACCESS reading, ACCESS writing, and ACCESS composite score), using data of the full sample

One should note that the hazard and survival rates are only available for listening and speaking during the first three years, from time 0 through time 2. This is because the MDE only required minimum proficiency scores for speaking and listening during a three-year period from 2013-2014 through 2015-2016. Those requirements were removed in 2016-2017. Given this

change in policy, students entering school before 2016-2017 and those entering school after that year were faced with different proficiency criteria. The three earlier cohorts (i.e., cohorts entering school prior to 2016-2017) were expected to meet (a) all five criteria during the first three years after they entered school (time 0 through time 2), and (b) only the criteria for reading, writing, and composite score from the fourth through the sixth years (time 3 through time 5). In contrast, the three later cohorts (i.e., cohorts entering school in or after 2016-2017) were only expected to meet the criteria for reading, writing, and composite score, but not those for speaking and listening. Therefore, the risk sets for speaking and listening were only comprised of students in the three later cohorts (but not those in the earlier cohorts).

Model estimated cumulative proficiency rates (listed in the last column of Table 28, visualized in Figure 16) indicate that students met the speaking and listening criteria rather early. Around 60% of students were already proficient in listening when finishing the first year of school; and in their third year, when most students were in second grade, over 90% of students attained a proficient score in listening. Speaking trailed closely behind listening in cumulative proficiency rates. While only 35% of students spoke English proficiently when they finished the first year, nearly 80% of students reached a proficient score in speaking by the end of the third year. The median time to proficiency was less than a year in listening and was around one year and a half in speaking. The observation that speaking and listening criteria were relatively easy for ELs to achieve is consistent with previous studies (e.g., Thompson, 2017; Umansky & Reardon, 2014).

Compared with listening and speaking, it took students longer to meet reading and writing criteria, or criteria that require literacy skills. Reading was a relatively small literacy-based barrier to proficiency for students in Michigan public schools. As shown in Table 28 (visualized

in Figure 16), nearly 70% of students scored proficient in reading in the third year, and the proportion increased to about 88% in the sixth year. The full sample's median time to reading proficiency was around two years, which is two years shorter than what was observed in previous studies (e.g., Thompson, 2017; Umansky & Reardon, 2014). Writing, however, posed a much larger barrier for students in Michigan public schools. By the end of the third year, the cumulative proficiency rate in writing was merely 15%, lagging behind the proficiency rate in reading by almost 55%. Although the pace at which new students met the writing criterion picked up starting in the fourth year, around 20% of students still had not reached proficiency in writing by the end of the sixth year. The median time to proficiency in writing (4.5 years) was the longest among the four domains.

It is somewhat surprising that writing was a larger barrier for students than was the composite score in every time period through the fourth year (as evidenced by the relative positions of the lines pertaining to ACCESS writing and ACCESS composite score in Figure 16). This was true despite the fact that the cut scores for writing were either equal to or lower than the cut scores for composite scores over the years of the study. Beginning in the fifth year, the composite score became a larger barrier to proficiency than writing, although by a relatively small margin. The median time to proficiency was slightly longer for the composite score (5 years) than for writing (4.5 years).

Findings from this study corroborate prior reclassification research suggesting that literacy-based language skills take longer to develop than speaking and listening skills (e.g., Umansky & Reardon, 2014; Thompson, 2017). However, unlike the previous research, this study identified writing, rather than reading, as the major literacy-based barrier to proficiency.

Chapter summary:

- Half of students who entered Michigan public schools in kindergarten as ELs attained proficiency (as measured by ACCESS against the criteria established by the MDE) in five years. About 35% of students were still not proficient in English even after spending six years in Michigan public schools.
- ELs with specific learning disabilities, ELs with speech/language impairments, and ELs with other types of disabilities were less likely to attain proficiency than their peers without disabilities.
- ELs speaking some particular home languages (e.g., Arabic, AfroAsiatic, Dravidian, Spanish) were less likely to attain proficiency than those who spoke some other languages at home (e.g., Bengali, Chinese, Japanese/Korean).
- ELs who ever lived in poverty were less likely to attain proficiency than their peers who never lived in poverty in each time period.
- ELs who ever used English at home were slightly more likely to attain proficiency in early elementary grades but were surpassed by their peers who never used English at home in late elementary grades by a small margin.
- ELs instructed bilingually, those instructed in English only, and those with no program were equally likely to attain proficiency throughout the six years of elementary schooling; there was suggestive but inconclusive evidence that ELs in dual-language bilingual programs surpassed those instructed in English only by the end of elementary school.
- ELs who were ever retained were more likely to attain proficiency than their peers who were never retained in early elementary grades, but this trend reversed in late elementary grades.

- Among the five individual proficiency criteria set by the MDE (i.e., reading, writing, listening, speaking, and composite score on ACCESS), writing appeared to be the largest barrier to proficiency for ELs in Michigan public schools.

## CHAPTER 9 DISCUSSION

Increased global migration and immigration has resulted in the growing presence of young language minority students in the United States and many other countries. In U.S. public schools, language minority students who are not proficient in English are commonly referred to as English learners (ELs). Young EL students are faced with a challenging task of learning academic content in English while developing their language proficiency in English as well as their first language (L1). Current U.S. education policies emphasize the education of ELs by holding state and local educational agencies accountable for this subgroup's language and academic outcomes; however, those policies are not always based on research or empirical data (Thompson, 2017). One federal mandate that is particularly relevant to this study is that states must establish yearly targets for ELs' English proficiency development. In the real world, states often struggle to establish reasonable targets due to two reasons: first, there is only limited research on child second language acquisition (SLA) and factors related to variation in child SLA compared to research on tertiary students and adults (Paradis, 2008); second, high-quality EL assessment data have only become available in relatively recent years, thus limiting researchers' abilities to examine EL children's long-term English proficiency trajectories (Kieffer & Parker, 2016).

This study aims to contribute to solving this issue by examining the time that ELs need to attain proficiency in English drawing on data from ELs in Michigan. English proficiency was measured by Michigan's state English language proficiency (ELP) assessment, ACCESS for ELLs (hereafter ACCESS), and was evaluated against the ELP standards established by the Michigan Department of Education (MDE) for the study period of 2013-2014 through 2018-2019. This study contributes to the existing literature in that it examines test-based English



language acquisition patterns in ELs independently from reclassification considerations that are irrelevant to the core construct of ELP (e.g., content-area assessments; attendance-record considerations). This study also extends previous work in the area by examining a variety of child, family, and school predictors of English learning outcomes simultaneously.

## **9.1 Time to Proficiency**

This study confirms prior research findings that it takes most ELs many years to become proficient in English (Hakuta et al., 2000). Specifically, I found that it took five years for around 50% of students who entered Michigan public schools as ELs in kindergarten to attain proficiency. Around 35% of those students were still not proficient in English after spending six years in school and were very likely to become long-term ELs (LTEs), a categorization that is starting to gain more attention in the applied linguistics and applied language education literature. The median time to proficiency obtained in this study (five years) is one to two years longer than time-to-reclassification estimates reported by researchers in New York City (Kieffer & Parker), Massachusetts (Slama, 2014), Texas (Slama et al., 2017), and Washington (Greenberg Motamedi et al., 2016), and is at least one year shorter than those estimates reported by researchers in California (Thompson, 2017; Umansky & Reardon, 2014). One explanation for these differences may be the student population studied. Unlike some previous studies that focused merely or predominantly on low-income ELs (e.g., Kim et al., 2014; Greenberg Motamedi et al., 2016) or Spanish-speaking ELs (e.g., Umansky & Reardon, 2014; Slama et al., 2017), this study used data from a more economically and linguistically diverse EL population representing over 200 home language groups. The variation in these estimates could also be explained by differences in proficiency/reclassification criteria used by states. For example, researchers in California have reported the longest time-to-reclassification estimates (up to eight

years) perhaps because reclassification in California involves several challenging criteria based on content-area assessments (Thompson, 2017; Umansky & Reardon, 2014). Comparatively, researchers in states that only use an ELP assessment for reclassification decisions have generally found shorter times to reclassification (e.g., New York; see Chapter 4 for a more thorough description). It should be noted that none of the previous studies used data from a WIDA state but were all located in non-consortia, “stand-alone” states. The definitions and measures of English proficiency in stand-alone states may reflect those states’ idiosyncratic understandings of ELP and may be different from the common understanding developed jointly by members of a large consortium like WIDA (Linguanti, Cook, Bailey, & MacDonald, 2016). Compared with those previous studies, this study achieves broader generalizability in that its time-to-proficiency estimates can be directly compared to similar estimates from the other 39 WIDA states (see more about generalizability in section 9.4 of this chapter). Given that Michigan tends to set the ACCESS cut scores at a lower level than other states within the WIDA consortium (see Linguanti & Cook, 2015, for a review of the cut scores set by different WIDA states for ACCESS during the year of 2015-2016), estimates from this study can be considered as lower bounds for the time required by ELs to attain proficiency in WIDA states.

Analysis of ELs’ time to proficiency revealed uneven proficiency attainment rates across years of school instruction. The rate at which students attained proficiency appeared to be rather slow (e.g., less than 10% of students attained proficiency each year during the first three years) until it reached a peak in the fifth and sixth years (i.e., over 30% of students attained proficiency each year), when the vast majority of students were in fourth and fifth grades. These findings support the MDE’s recommendation against reclassifying ELs before the third grade (MDE, 2020). The observation of a peak in students’ likelihood of attaining proficiency toward the end

of elementary school is in line with previous studies that have identified an important reclassification window during upper elementary grades (Thompson, 2017; Umansky & Reardon, 2014). These studies also suggest that the epicenter-like, upper-elementary-school reclassification window may be ephemeral, because proficiency attainment may slow down in middle school as proficient students are more likely to be reclassified as former ELs prior to reaching the secondary level (Umansky & Reardon, 2014). Future research should follow ELs for a longer period of time to see how their proficiency attainment rates change from the elementary years through the middle and high school years. A difficulty in doing this now is the lack of 13-year, longitudinal datasets, which would allow for this type of work. The current COVID-19 pandemic hampers or puts a strong speed bump into such efforts, as it has, world-wide, either broken or punched serious holes in education and the longitudinal measurement of children's educational growth.

## **9.2 Factors that Affect Time to Proficiency**

### ***9.2.1 Primary Disability Type***

Results of the conditional hazard model showed that ELs' likelihood of attaining proficiency varied significantly by primary disability type. Among students in the subsample, those without disabilities were about two times more likely to attain proficiency than their peers with speech or language impairments and were about 12 times more likely to do so than their peers with specific learning disabilities. Accordingly, the median time to proficiency was estimated to be shortest for students without disabilities (4 years); the estimate was slightly longer for students with speech or language impairments (4.5 years), and much longer for students with specific learning disabilities (more than 6 years; exact estimate could not be obtained because students were not tracked long enough). Those patterns in general corroborate

Kieffer and Parker's (2016) observation of the relationship between disability and time to reclassification. Drawing on the data from ELs in New York City public schools, Kieffer and Parker found that the median time to reclassification was approximately 3.5 years for ELs without disabilities, 6 years for students with speech or language impairments, and 8 years for students with specific learning disabilities. Differences between these two studies' estimates can be explained in part by different events under investigation: while Kieffer and Parker studied reclassification, an event involving teachers' or school administrators' subjective judgments, this present study focused on proficiency attainment, an event objectively determined by students' ELP test scores. Depending on how local educators use ELP assessments and other criteria to make reclassification decisions for ELs with disabilities (see Schissel & Kangas, 2018 for a review), time to proficiency may be longer or shorter than time to reclassification for this population. Nonetheless, this present study and Kieffer and Parker (2016) have both found that ELs with disabilities, particularly those with specific learning disabilities, take substantially more time to attain proficiency in English or become reclassified, compared with their peers without disabilities. Disability is also found to be a factor contributing to increased likelihood of becoming LTELs. For example, data of the full sample indicated that over 90% of ELs with specific learning disabilities were still not proficient in English after six years and were very likely to become LTELs (also see Kieffer & Parker, 2016).

These findings suggest that ELs with disabilities may experience significant difficulties in learning English, pointing to the need for providing both language-learning and special-education services to those dually identified ELs. Indeed, research on dually identified ELs is newly growing, due to ELs being mandated to be classified, for the first time, as both ELs and disabled learners starting in 2015 under ESSA (ESSA, 2016,

<https://www2.ed.gov/policy/elsec/leg/essa/essatitleiiiguidenglishlearners92016.pdf>): Before the mandate, states, districts, or schools could decide to classify a student as an EL or as disabled, and could decide not to report the student as both. Much of the recent research on disabled ELs has shown that this population is indeed poorly served at school. Compared to ELs without disabilities, ELs with disabilities are more likely to receive few language-learning services and to be instructed only in English (Kangas, 2014; 2018; Zehler, Fleischman, Hopstock, Pendzick, Stephenson, 2003). Moreover, they tend to receive special education in segregated contexts (Zehler et al., 2003). In middle and high schools, ELs with disabilities are more likely to be placed in lower academic tracks, often with constrained access to high-quality curriculum and instruction (Kangas & Cook, 2020). All these factors make ELs with disabilities a particularly vulnerable population at school, one which is also historically understudied due to the group being historically under-identified, factors that may have contributed to their prolonged EL status and academic underachievement. Findings from this study underline that Michigan public schools should perhaps provide more attentive services to ELs with disabilities, and could start by diagnosing and documenting EL disabilities with more deliberateness. Importantly, EL and special-education educators should attend to the heterogeneity within this population and provide instructional interventions that are differentiated by ELs' specific types of primary (and preferably secondary) disabilities.

### ***9.2.2 Primary Home Language***

In this study, I identified and compared 13 home language groups' English acquisition patterns, with Spanish ELs serving as the reference group. In contrast to the highly cited finding that Spanish ELs are less likely to attain proficiency than non-Spanish ELs (e.g., Slama, 2014; Thompon, 2017), I found that the effect of home language on the likelihood of proficiency

attainment was dependent on years of school instruction. Although Spanish ELs acquired English at a slower rate than most non-reference groups (i.e., Arabic, Albanian, Austronesian, Bengali, Chinese, Dravidian, IndoIranian, Italic/Germaic, Japanese/Korean, and other home languages) in early elementary years, Spanish ELs were able to catch up and even surpass several non-Spanish groups over time. For example, Spanish ELs were expected to have a slightly higher cumulative proficiency rate than Arabic ELs by the end of the sixth year, despite the fact Spanish ELs were 1.50 times less likely to attain proficiency than Arabic ELs during the first year. Overall, Spanish ELs ranked middle to low among the 13 home language groups in terms of rate of English acquisition. Using Spanish ELs as the baseline, the 12 non-reference groups could be roughly classified into three categories based on their cumulative proficiency rates in the sixth year: those performed better than Spanish learners (Austronesian, Bengali, Chinese, and Japanese/Korean ELs), those performed worse than Spanish learners (Arabic, AfroAsiatic, and Dravidian ELs), and those performed comparably to Spanish learners (Albanian, BaltoSlavic, Italic/Germanic ELs).

These findings highlight substantial heterogeneity in the rate of English learning among ELs from different home language backgrounds. It is not completely clear, however, why home language relates significantly to the likelihood of proficiency attainment, even after controlling for other variables such as instructional programming, race/ethnicity, and poverty status. Previous research suggests that the mechanism governing the relationship between home language and English learning may be rather complex. A plausible explanation is the linguistic distance between English and ELs' primary home language (see Paradis, 2011); Conger (2010), however, argued that linguistic distance is unlikely the only explanation. For example, Arabic and Chinese are both considered to be linguistically more distant from English than Spanish is in

terms of the lexicon, syntax, and writing system (Gor & Vatz, 2009; Stevens, 2006). The implication is that English should be more difficult to learn for Arabic and Chinese speakers than for Spanish speakers. Indeed, studies of American college students learning foreign languages find that they take substantially longer to develop proficiency in less commonly taught languages (e.g., Chinese) than they do in Spanish (e.g., Zhang, Winke, & Clark, 2020), although a nod must be given to the notion that adults may be able to learn certain linguistic structures at a faster pace than children can due to adults' higher levels of cognitive maturity that allows them to tap into explicit learning contexts. Adults may also venture into higher-level academic and literacy contexts in instructed settings in a shorter amount of time than children do, due to adults' ability to transfer knowledge from their home-language literacy skills (and especially when their L1 literacy skills are high). With regard to this study's findings, although linguistic distance can be used to account for Spanish ELs' higher cumulative proficiency rates relative to Arabic ELs in upper elementary grades, it does not adequately explain why Chinese students acquired English faster than Spanish learners throughout the elementary school. Conger (2010) suggested that home language may act as a proxy for other unobserved traits that affect English acquisition, such as parental education. Umansky, Callahan, and Lee (2020) further suggested that researchers examine factors at multiple layers (micro-, meso-, and macro-levels factors) to understand the mechanism driving disparities in educational outcomes across different home language groups. Umansky et al., for example, found that individual background, social capital, school and instructional contexts, and racial or ethnic stereotypes and bias all contributed to differences between Chinese and Spanish ELs' reclassification patterns. As an important addition to the prior literature, this study challenges the prevalent belief that Spanish ELs are poorer English learners than non-Spanish ELs as well as the common practice among EL researchers of

operationalizing home language in an oversimplified, dichotomous manner as if ELs were either Spanish- or non-Spanish-speaking (e.g., Burke et al., 2016; Slama, 2014). Future researchers should carefully investigate differences in language and literacy outcomes across more sophisticatedly defined home language groups as well as factors associated with those differences.

### ***9.2.3 Poverty Status***

Poverty status, as indicated by an EL's eligibility for free/reduced-price lunch in a given year, was a significant time-varying predictor for proficiency attainment. ELs ever receiving free/reduced-price lunch were nearly two times less likely to attain proficiency than their peers never in poverty in every time period from the first through the sixth years of schooling. Compared with previous studies that also used survival analysis to examine the relationship between poverty and the likelihood of reclassification/proficiency, this study's estimate of poverty effect tends to be larger than prior studies that treated poverty as time-invariant (e.g., Thompson, 2017), and tends to be similar in magnitude to those studies that treated poverty as time-varying (e.g., Kim et al., 2014; although Burke et al., 2016 and Slama, 2014 performed similar analyses, they did not control for influential confounding variables such as initial proficiency in English). This suggests that treating poverty as time-invariant may mask important changes in ELs' long-term poverty patterns, and therefore underestimate the effect of poverty on ELs' educational outcomes. In an earlier study, Micheltore and Dynarski (2017) found that students suffering from chronic poverty experience substantially more difficulties in learning English reading than their peers who are occasionally poor or never in poverty. Their findings supported the idea of coding poverty as a time-varying variable to more faithfully reflect students' changing poverty patterns in longitudinal research. This present study and Kim et al.



(2014) both illustrate that it is easy to include poverty as a time-varying predictor in longitudinal data analyses (e.g., survival analysis, latent growth curve modeling).

Although the influence of poverty on academic outcomes has been well studied in general education, researchers in the field of SLA are still at an early stage in understanding how and why socioeconomic status (SES) affects a person's second language learning (Bulter, 2014). Butler and Le (2018) found that SES (defined in terms of family income and parental education) may impact SLA through important mediating variables including the amount of language-learning resources available and parental beliefs and expectations about children's success in learning the target language. As Butler and Le suggested, future research is necessary to explore the mechanisms governing the relationship between SES and SLA across time and contexts. In addition, more research is needed to inform the types of classroom, school, and community supports that are effective in helping ELs suffering from chronic poverty develop their English skills.

#### ***9.2.4 Home English Use***

This study confirms previous research findings that there is tendency for ELs to switch from using their home language (or first language, L1) to using English at home after they start school (Kim et al., 2020; Mancilla-Martinez & Lesaux, 2011). Additionally, I found considerable variation in ELs' home English experience with regard to whether and when they switched to using English at home. Data of the 2013-2014 kindergarten cohort indicated that ELs who experienced a home language shift most likely did so in their second grade (Year 3 in the dataset) and tended to continue using English in the home thereafter.

A question of great interest to educators, policymakers, and perhaps many parents is whether using English at home facilitates ELs' English acquisition. This study found that ELs

who *ever* used English at home were slightly more likely to attain proficiency than their peers who *never* used English at home in early elementary grades, but this trend reversed over time, with those who *never* used English at home becoming more likely to attain proficiency in upper elementary grades. In the sixth year (i.e., when most students were in fifth grade), students who *never* used English at home were expected to have a slightly higher cumulative proficiency rate than those who *ever* used English at home. Although these effects were small, they demonstrated statistical significance and connect in several important ways with prior literature on the role of home language environment on ELs' language and literacy development. First, this study provides little support for the time-on-task hypothesis that posits a direct relationship between the amount of exposure in a language (i.e., English) and the likelihood of proficiency in that language (Mancilla-Martinez & Lesaux, 2011). As Pearson (2007) suggested, this is perhaps because the time-on-task hypothesis is constrained by a boundary condition; that is, once the amount of exposure reaches a certain threshold, or so-called "critical mass," a direct relationship between exposure and learning disappears (i.e., more exposure does not matter). The evidence that the effect of ever using English at home was strongest in early elementary grades and declined over time suggests that a threshold may truly exist and may be reached as students' exposure to English accumulates at school. Without invoking the threshold hypothesis, Duursma et al. (2007) echoed this point and argued that ELs' exposure to English at school is so powerful that additional exposure to English at home is not necessary for them to attain proficiency. Second, results of this study provide some evidence that using the L1 rather than English at home during the school years may be more beneficial for ELs' English development in the long term. This is consistent with the findings from a recent longitudinal study in Canada (Kim et al., 2020). Kim et al. compared the literacy achievement patterns of four groups of ELs living in

different home language environments (i.e., switching from L1 to English, switching from English to L1, always using L1, always using English) and found that the group that never lived in an English-dominant home (i.e., always using L1) between grades 3 and 6 had the most favorable reading outcomes from grade 3 through grade 10. Although this present study only found a modest advantage associated with the usage of L1 relative to English, the fact that such findings are statistically significant could potential contribute to a large body of research on the broad benefits of bilingual competence (e.g., cognitive, socio-behavior, economic benefits; see Bialystok, 2001; 2011), and could, at a minimum, suggest that maintaining the L1 in the school years does not impede ELs' English acquisition (Duursma et al., 2007; Hammer et al., 2009; Kim et al., 2020). Third, an important question that I was not able to investigate in this study is the effect of home language environment on ELs' L1 proficiency development. Previous research suggests that although attaining English proficiency does not require the use of English at home, becoming or staying proficient in L1 may rely critically on L1 maintenance at home (Duursma et al., 2007; Gutiérrez-Clellen & Kreiter, 2003). Some studies have further shown that parents' increased English usage at home could be associated with a decline in children's L1 growth (Hammer et al., 2009) or even loss of the L1 (Wong Fillmore, 1991). Future researchers should measure EL outcomes in both the L1 and English to determine the contribution of home language environment to ELs' language development.

### ***9.2.5 Instructional Programming***

I investigated the differential effects of four types of language-learning programs on ELs' English language development: English as a second language (ESL), transitional bilingual, dual-language bilingual, and no program (participation in mainstream classes, with no ESL services). After controlling for other variables (e.g., initial proficiency in kindergarten), instructional

programming demonstrated a negligibly small effect on ELs' likelihood of attaining proficiency in each time period. The predicted time-to-proficiency patterns were similar across the four programming groups with some interesting but non-significant trends: (a) ESL students had slightly higher cumulative proficiency rates in early elementary grades, but students in bilingual programs closed the gaps over time, with those in dual-language bilingual programs even surpassing their ESL peers in upper elementary grades by a small margin; and (b) students with no program were slightly more likely to attain proficiency than those participating in language-learning programs. These findings partially support previous studies (Slama et al., 2017; Thompson, 2017; Umansky & Reardon, 2014). For example, Umansky and Reardon (2014) similarly found that students in bilingual programs had more favorable educational outcomes in the long run despite their slower reclassification rates in elementary grades. However, unlike this study, Umansky and Reardon found those effects to be statistically significant, perhaps because they tracked the ELs for a longer period of time through high school years. Researchers have noted that the benefits of bilingual programs may only appear in the long term, perhaps past the elementary grades (Robinson-Cimpian, Thompson, & Umansky, 2016; Umansky & Reardon, 2014). With regard to the effect of participating no program on English acquisition, Slama et al. (2017) found that students with no program were less likely to attain proficiency than those in bilingual programs, contrary to the finding of this study. However, one should avoid overinterpreting these patterns because neither this study nor Slama et al. reported results that were statistically significant nor practically meaningful.

In sum, findings of this study indicated that ELs in bilingual programs performed as well as their peers in ESL programs in acquiring English in elementary school. In other words, providing ELs with opportunities to develop their native language and literacy skills while they

are learning English at school does not harm or interfere with ELs' English acquisition. These findings add to a growing body of research that contradicts the time-on-task hypothesis (e.g., Rossell & Baker, 1996) that associates faster rate of acquisition with more exposure to English. Previous studies comparing English literacy outcomes between ELs instructed in ESL and bilingual programs have either found no differences or an advantage for ELs instructed bilingually (Slavin, Madden, Calderón, Chamberlain, & Hennessy, 2011; Steele, Slater, Miller, Zamarro, & Li, 2017; see Slavin & Cheung, 2005 for a meta-analysis). In addition, ELs in bilingual programs have been found to significantly outperform their ESL peers on assessments given in the partner languages used in bilingual programs (Barnett, Yarosz, Thomas, Jung, & Blanco, 2007). Collectively, existing research has suggested that providing bilingual education to ELs may bring more benefits to those students than instructing them only in English, and such benefits may become increasingly apparent in the long term (Robinson-Cimpian et al., 2016). However, more research is still needed to determine the exact characteristics of bilingual programs that may have contributed to more favorable long-term EL outcomes relative to ESL programs (Umansky & Reardon, 2014). While *language* of instruction could be an important contributor, some researchers have noted that *quality* of instruction may be the ultimate determiner in EL education (Slavin & Cheung, 2012). Every person probably remembers a particular teacher who impacted their educational trajectory the most: A teacher under whom the person thrived. Educational data, such as those used in this study, thus not surprisingly do not capture the effects of particularly effective teachers or particularly effective EL programs. To obtain a more robust understanding of the effect of instructional programming on ELs' educational outcomes, future research should carefully examine how students are placed into different programs and how individual programs differ from each other in classroom

implementation. Qualitative research in educational linguistics or mixed-method studies may indeed be extremely promising avenues of research to uncover the impacts of high- versus lower-quality EL programs and classrooms.

#### **9.2.6 Retention**

In this study, I investigated the relationship between traditional retention (as opposed to test-based retention) and ELs' likelihood of proficiency attainment. Around 8% of ELs in the full sample were ever retained due to traditional policies. This rate is slightly lower than the retention rate for students at the national level (10%; Planty et al., 2009) and is substantially lower than the retention rates previously reported for ELs in other states (e.g., Slama, 2014 reported a traditional retention rate of 21.6% for ELs in Massachusetts; Figlio and Özek, 2020 reported a test-based retention rate of 18.5% for ELs in Florida). The data of the 2013-2014 kindergarten cohort suggested that students who ever experienced retention were most likely retained in kindergarten. Although ELs who were ever retained tended to achieve lower average ACCESS scores than those who were never retained, it is unclear if low English proficiency was one of the reasons why the students were held back.

Results of the conditional hazard model indicated that retention (operationalized as whether a student was ever retained in a given year) had a time-varying effect on students' odds of attaining proficiency. Specifically, the effect of retention was moderate and positive in the second year of school (i.e., the first time when students could be retained), with students ever retained being about 1.5 times more likely to attain proficiency than their peers who were never retained; however, the positive effect declined rapidly over time and eventually became moderate and negative in the sixth year. This finding is consistent with previous studies that find that transitional retention is associated with short-term gains followed by increasingly negative

impacts on students' academic performance in the long run (see Allen et al., 2009 for a meta-analysis). In contrast, more recent research examining the effect of test-based retention has provided overwhelming positive evidence for the efficacy of retention, at least on test scores (e.g., Figlio and Özek, 2020). Figlio and Özek (2020), for example, reported a variety of test-score-based benefits associated with test-based retention for ELs in Florida, including reducing the median time to English proficiency by half (social and family effects were not part of the study). The fact that Figlio and Özek (2020) focused on test-score-based (sometimes called “automatic” retention) rather than traditional (educator-decided) retention may in part explain why they found different results from this present study. In particular, Florida’s test-based retention policy, which has been in place since 2002, required all retained ELs to attend a summer reading program prior to the next school year. Because Figlio and Özek were not able to disentangle the effect of retention from that of the summer program, their estimated retention effect may have been inflated by the additional instruction that retained students received during summer, which also seemed to focus on reading-test practice. Methodological differences could also be a possible explanation. While this study employed regular logistic regression to investigate the effect of retention, Figlio and Özek used a quasi-experimental method called regression discontinuity, which allowed the researchers to examine the causal relationship between retention and student-test-score outcomes based on pre-determined retention criteria (i.e., designated cut scores on content-area assessments). Causal inferences, however, could not be drawn from this study due to unobserved characteristics that may have affected retention decisions.

### 9.3 Barrier to Proficiency

The Michigan Department of Education (MDE) established five criteria for English proficiency during 2013-2014 through 2018-2019, including: (a) reaching a designated minimum composite score on ACCESS (2013-2019); (b) reaching designated cut scores on the reading and writing domains of ACCESS (2013-2019); (c) reaching designated cut scores on the speaking and listening domains of ACCESS (2013-2016). Separate analyses of the five individual criteria showed that students met speaking and listening criteria relatively quickly, whereas achieving literacy-based criteria (i.e., reading and writing) required longer times. This is consistent with previous studies that examined time to proficiency in different language skills (Hakuta et al., 2000; Thompson, 2017; Umansky & Reardon, 2014). However, readers should note that this study's time-to-proficiency estimates for speaking and listening were only based on three years of data (2013-2016) from students in the three older cohorts (i.e., cohorts that entered school prior to 2016-2017). This is because the MDE dropped the requirements for speaking and listening in 2016-2017, the year when WIDA conducted a standard-setting study and raised the proficiency standards for ACCESS (Cook & MacGregor, n.d.). Descriptive statistics (reported in Chapter 7) suggest that students' average ACCESS scores declined to some extent in all four domains in and after 2016-2017, with speaking being influenced the most. (Readers should note that this does *not* mean that students' skills declined; rather that students were evaluated against harder or different proficiency standards.) It is therefore unclear whether the results would remain the same if the MDE had kept the criteria for speaking and listening between the years of 2016-2019.

Writing was found to be the largest barrier to proficiency among the five individual criteria. For example, the cumulative proficiency rate in writing was merely 15% by the end of



the third year, whereas the rate was 93% in listening, 79% in speaking, 69% in reading, and 36% in overall proficiency (as measured by the ACCESS composite score) in the same year. This is a somewhat surprising finding given that the cut scores for writing were either equal to or lower than the cut scores for other domains and the composite score during the period of 2013-2019. This finding contrasts with the results of previous studies that suggest reading as the most difficult skill for ELs (Thompson, 2017; Umansky & Reardon, 2014). The lower likelihood of proficiency in writing may reflect the difficult nature of acquiring specific aspects of English writing (e.g., transcription skills, oral language, etc.; Williams & Lowrance-Faulhaber, 2018). It may also reflect potential problems with the writing domain of ACCESS (an unfriendly test design, inappropriate scoring rubrics, or overly high proficiency cut scores) or with the curriculum, instruction, and support that ELs receive at school for writing development. Furthermore, this finding may raise questions about the alignment between the external writing assessment (i.e., ACCESS) and classroom-based curriculum and writing instruction for ELs. It could be that in the classroom, students' writing, especially in the early elementary grades, is only considered formative, and not something to be evaluated in a summative manner, which would allude to a disconnect between classroom-based writing and literacy-achievement objectives and theories, and the hypotheses upon which child-writing tests rely for accurate and valid score interpretation. More research is needed to understand why writing represented a larger barrier than other English skills for ELs in Michigan public schools.

#### **9.4 Generalizability**

Generalizability has been mentioned multiple times throughout this dissertation. Here, I discuss the generalizability of this study's results by relating the results to previous research and by discussing this study's real-world constraints. Compared with previous researchers who

examined time to reclassification, I achieved broader generalizability with this study by examining time to English proficiency. As I explained in Chapter 4, time-to-reclassification estimates are dependent on a state's (or district's) specific reclassification policy (e.g., English proficiency assessment, content-area assessments, test-based proficiency standards, non-test-based reclassification requirements) as well as local educators' interpretation of the state's (or district's) policy. As such, time-to-reclassification estimates are usually not generalizable beyond the state (or district) in which those estimates are obtained. Time-to-proficiency estimates, however, are less dependent on the context under investigation. Although time to proficiency is also tied to a state's (or district's) choice of English proficiency assessments and the associated proficiency standards, this estimate is not influenced by reclassification requirements that are irrelevant to the construct of English proficiency, such as content-area assessments, nor is it influenced by local educators' subjective judgments about an EL student's English ability or readiness to join the mainstream classroom. Due to these reasons, time-to-proficiency estimates are more likely to generalize across states than time-to-reclassification estimates, particularly within assessment consortia where member states use the same English language proficiency assessment and proficiency standards.

A second factor that contributes to this study's broader generalizability is the choice of the study context of Michigan, which is a member of a large, multi-state English assessment consortium, WIDA. Within the WIDA consortium, there is a consensus among states (including Michigan) on using WIDA's English Language Development Standards as their English proficiency standards and ACCESS as their English proficiency assessment. Situating this study in a WIDA state thus enhances its generalizability at the national level in that direct comparisons may be drawn between this study's time-to-proficiency estimates and similar estimates obtained

in the other 39 WIDA member states. To my best knowledge, this is one of the first studies to investigate time to proficiency within the WIDA consortium. Previous research was predominantly located in non-consortia, “stand-alone” states (California, New York, and Texas), thus having low generalizability at the national level (see Chapter 4 for a review).

Despite the steps that I took to ensure the generalizability of this study’s results, one should keep in mind that comparisons of time-to-proficiency estimates across WIDA states are constrained by the variation in the types of ACCESS criteria (e.g., composite score, domain scores) and the associated cut scores that WIDA states use to benchmark English proficiency (see Linquanti and Cook’s work, 2015, for a summary of the proficiency criteria and cut scores used by WIDA states in 2015-2016). The fact that Michigan and perhaps many other WIDA states adjusted their proficiency criteria and cut scores from time to time could limit the generalizability of this study’s results. (However, as I will explain in Chapter 11, those adjustments were made to improve the validity of the interpretations and uses of the ACCESS test scores and thus had benefits that far outweigh researchers’ interests in broad generalizability.) Caution should thus be exercised when one compares this study’s estimates to those from a WIDA state that used different criteria or cut scores from Michigan during the study period.

## **9.5 Is Survival Good or Bad?**

Lastly, I would like to briefly revisit issues with using the traditional survival analysis terminology, which was first developed by medical researchers for death events, to describe time-to-event analyses in non-medical fields like social science, general education, and applied linguistics (see Chapter 7 for a detailed description of the issues). When I first learned about survival analysis, it took me a while to shift my mindset and accept that survival does not

necessarily mean something good, and hazard and risk do not necessarily indicate something bad. In fact, in social science, general education, and applied linguistics, “surviving” an event is often a bad thing, whereas having a high “risk” of experiencing the event is a good thing. This is because many events in these fields are positive in nature, such as the event under investigation in this study, proficiency attainment. Using negative terms to describe a positive event (e.g., some student has a risk of attaining proficiency) may thus create confusion for readers who are not familiar with survival analysis and its traditional terminology. In those cases where the event of interest is neutral, careful selection of terminology may be even more important. For example, many researchers are interested in the likelihood of ELs becoming LTELs (e.g., Kieffer & Parker, 2016). While the event of becoming an LTEL itself is neutral, the way some researchers describe this event—for example, “students at risk of becoming LTELs” (Olson, 2014, p. 22)—may make readers consider LTELs are linguistically less competent than ELs who reclassify in less than six years, despite the fact that LTELs may have stronger bilingual competence than their English-proficient peers (Flores & Rosa, 2015). Although terms like “risk” and “hazard” are neutral within the statistical context of survival analysis, readers may take those terms out of context and interpret them as something negative. A key implication is that researchers in non-medical fields should adjust the traditional survival analysis terminology so as to fit their specific research contexts. In Table 29, I provide some adjusted terms that I presented earlier in Chapter 7. I hope that this table can inspire researchers in social science, general education, and applied linguistics to find terms that more appropriately describe their analyses.

Table 29. Traditional and adjusted survival analysis terminology for social science, general education, and applied linguistics

Traditional statistical terms	Adjusted terminology for social science	Adjusted terminology for general education/ applied linguistics/this study
Survival analysis	Event history analysis	Proficiency attainment analysis
Risk	Likelihood of experiencing the event	Likelihood of proficiency attainment
Risk set	Likelihood set	Likelihood set
Hazard	Conditional event probability	Conditional attainment probability
Survival probability	Cumulative event probability	Cumulative attainment probability
Censored	Participants with unknown event time	Learners with unknown attainment time

#### Chapter summary:

- This study confirms prior research findings that it takes most ELs many years to become proficient in English (Hakuta et al., 2000) and that the likelihood of proficiency attainment peaks for elementary-school ELs when they reach late elementary grades (Thomspon, 2017).
- Consistent with previous literature on ELs with disabilities (Kieffer & Parker, 2016), this study found that ELs with specific learning disabilities and those with speech/language impairments were significant less likely to attain proficiency than their peers without disabilities throughout elementary schooling. These findings suggest that dual services (language learning and special education) should be provided to ELs with disabilities to help them overcome difficulties in learning English.
- In contrast to many previous studies that suggest that Spanish ELs are less likely to become proficient than non-Spanish ELs (e.g., Slama, 2014), this study found that Spanish ELs acquired English at a faster rate than ELs from some non-Spanish L1

backgrounds (e.g., Arabic and Dravidian ELs) and performed comparably to ELs from some other L1 backgrounds (Albanian, BaltoSlavic, Italic/Germanic ELs). These findings challenge the common practice among EL researchers of operationalizing home language in an oversimplified, dichotomous manner as if ELs were either Spanish- or non-Spanish-speaking (e.g., Burke et al., 2016; Slama, 2014)

- In line with some prior research (Duursma et al., 2007; Hammer et al., 2009; Kim et al., 2020; Umansky & Reardon, 2014), this study found that ELs who received partial instruction in the L1 and ELs who never used English at home were equally likely, and in some cases slightly more likely, to attain proficiency as compared with their peers who were immersed full-time in an English-only environment at school and at home. These findings highlight the need for more research on the role of L1 maintenance in English proficiency development.
- ELs who ever lived in poverty were less likely to attain proficiency than those who never lived in poverty, consistent with previous findings (Kim et al., 2014; Thompson, 2017). Findings also suggest that researchers should improve their methods of coding poverty status to more faithfully reflect students' changing poverty patterns in longitudinal research.
- Consistent with previous literature on the effect of traditional retention on ELs (Allen et al., 2009), this study found that retention led to more negative than positive impacts on ELs' English proficiency outcomes in the long term. There is a need to investigate why test-based retention (often coupled with additional

instructional supports) appears to be associated with more positive effects than traditional retention.

- Unlike previous studies that identified reading as the largest barrier to proficiency for ELs (e.g., Thompson, 2017), this study found that it took ELs the longest to attain proficiency in writing. More research is needed to uncover why writing was the most difficult English skill for ELs in Michigan.
- Traditional survival analysis terminology carries a negative connotation. There is a need for using adjusted terminology to describe survival analysis in the fields of social science, general education, and applied linguistics.

## CHAPTER 10 PRACTICAL IMPLICATIONS

This study has several practical implications for policymakers, educators, test developers, and parents.

### 10.1 Improving School Accountability Systems to Reflect the Diversity of the EL Subgroup

Findings of this study corroborate prior research that indicates that ELs are a highly heterogeneous population (National Academies of Science, Engineering, and Medicine, 2017). Researchers have long argued that school accountability systems should be improved to reflect the diversity of the EL subgroup (e.g., Burke et al., 2016; Cook et al., 2012; Slama et al., 2017; Thompson, 2017). Texas, for example, has established differentiated growth expectations for ELs by taking account of initial English language proficiency and time in U.S. schools (Slama et al., 2017). Supporting this act of Texas, results of this study additionally suggest that other EL characteristics (such as primary disability type, primary home language, and poverty) should also be considered when establishing educational targets for ELs.

First, this study found that primary disability status had a significant effect on ELs' abilities to attain proficiency. Particularly, it appeared to be extremely difficult for ELs with specific learning disabilities to score proficient on ACCESS (i.e., over 90% of ELs with specific learning disabilities were still not proficient on ACCESS after spending six years in Michigan public schools). This result is very concerning considering that ELs in Michigan must meet all proficiency criteria for ACCESS to become reclassified and gain access to the mainstream curriculum. In their investigation of reclassification policies in New Mexico and California, Schissel and Kangas (2018) expressed similar concerns about policies and practices that made reclassification improbable for ELs with disabilities. Schissel and Kangas warned that the unlikelihood of being reclassified has profound educational consequences for ELs with



disabilities (e.g., becoming long-term ELs), putting them at elevated risk for dropping out of high school before graduation. Considering these research findings, the Michigan Department of Education (and perhaps policymakers in other WIDA states) may want to re-evaluate whether ACCESS for ELLs (here after ACCESS) is an appropriate and valid English proficiency assessment for ELs with disabilities. An option is to allow all ELs with disabilities to take Alternate ACCESS for ELLs (or Alternate ACCESS, a large-print, paper-based test individually administered to students in Grade 1-12 who are identified as ELs with the most significant cognitive disabilities) and to establish separate reclassification criteria for ELs with disabilities that do not depend completely on their standardized test scores. Policymakers and the developers of ACCESS should also collaborate to improve the types of accommodations ELs with disabilities receive when taking ACCESS, making sure that the accommodations are research-based, effective, and valid (Schissel & Kangas, 2018). Most importantly, EL educators and teachers must provide the dual services (i.e., special education and specialized language services) that ELs with disabilities need to attain proficiency in English and content areas. This entails expanding opportunities to learn for these students throughout all junctures of their education (Kangas & Cook, 2020).

Second, this present investigation found that home language and socioeconomic backgrounds were significantly related to variation in ELs' time to proficiency, even after controlling for other factors such as instructional programming and initial English proficiency. Several researchers have pointed out that it is a serious problem when sociodemographic factors beyond the control of teachers and school administrators (e.g., home language, poverty) are better predictors of student performance on high-stakes accountability measures than factors pertaining to the quality of teaching and learning (Baker & Johnston, 2010; Tavassolie &

Winsler, 2019). The finding that ELs living in poverty and ELs speaking some particular home languages (e.g., Arabic, AfroAsiatic, Dravidian, and Spanish) had significantly lower likelihood of attaining proficiency indicates that ACCESS may assess home financial situations and/or linguistic and cultural differences (or other factors that are implicated in poverty and home language variables, e.g., parental education, attitudes towards integration into English-speaking) more so than it assesses student learning. This could call into question the validity of ACCESS in measuring ELs from different backgrounds. It also challenges the current accountability practices of establishing a single uniform standard of performance and holding all students accountable to reach such a standard. Test developers and policymakers should work to improve the test-based accountability system to ensure fairness for ELs from diverse backgrounds. Indeed, ELs are what many researchers are now calling a “superdiverse” group, meaning that this group is diverse not only in terms of ethnicity or country of origin—the variables that have been historically used to describe immigrants—but also a wide variety of other variables (gender, age, economic mobility, locality, sexuality etc.) that may interact to influence the life trajectories of ELs (Blackledge, Creese, & Takhi, 2013). Based on the findings of this study, it is important that schools provide extra support (e.g., tutoring, reading materials, summer instruction etc.) to ELs from low-income families or ELs from home languages associated with low odds of proficiency to ensure those children have more equal opportunities for success in school.

## **10.2 Taking More Measures to Support ELs’ L1 Development**

Contrary to the prediction of the time-on-task hypothesis (Rossell & Baker, 1996), I did not find that immersing ELs in a full-time English-only environment at home or at school accelerates their English acquisition. Instead, results of this study agree with previous findings suggesting that exposing ELs to their home language (L1) during the school years (whether at

home or at school) does not hinder or interfere with their English learning (Duursma et al., 2007; Kim et al., 2020). In addition, there is suggestive but inconclusive evidence indicating that maintaining the L1 in the home and school contexts may be more beneficial for ELs' English proficiency outcomes in the long run.

These findings have important implications for L1 maintenance. Many educators and parents believe that exposing children to English in the home would help them acquire English at a faster pace (Hammer et al., 2009; Paradis, 2011). This belief, however, is not supported by this study or much of the previous literature (Duursma et al., 2007; Gutiérrez-Clellen & Kreiter, 2003; Paradis, 2011). Furthermore, some prior studies suggest that parental usage of English may lead to serious negative consequences for EL children, including the loss of the L1 (e.g., Wong Fillmore, 1991) and loss of a bilingual advantage (Bialystok, 2001). Maintaining the L1 in the home not only prevents such consequences, but it can also bring many benefits to ELs, such as healthier identity development, closer family relationships, and greater access to economic opportunity (Surrain, 2018). To harvest these benefits, parents need to support bilingualism and the use of the L1 at home (Hwang, Mancilla-Martinez, Flores, & McClain, 2019). Successful L1 maintenance is further dependent on educators' and policymakers' efforts to create school and sociopolitical environments that are safe and friendly for immigrants and their children and that promote EL bilingualism (Dixon & Wu, 2014; Surrain, 2018). One way to do so is to advocate for bilingual education. Over the past 30 years, states across the nation have been marginalizing bilingual programs under the influence of the English-only movement (Cheng & Slavin, 2012). In several states, it manifests via anti-bilingual legislation (e.g., California, Arizona, Massachusetts). Although Michigan does not legislate to curtail bilingual education, it is clear from the data of this study that English immersion is the dominant EL program model across the

state and that the number of bilingual programs has been declining rapidly over the recent years. Supporters of anti-bilingual actions often invoke the time-on-task hypothesis, arguing that L1 maintenance impedes learning of the school language (Cummins, 2019). These arguments, as I mentioned earlier, are not adequately supported by the existing literature (Cummins, 2019), nor by the data in this study. Indeed, a growing body of studies have found that bilingual programs, particularly those using a dual-language instruction model, afford more favorable English language and literacy outcomes for ELs in the long run as compared with English-only instruction (Robinson-Cimpian, Thompson, & Umansky, 2016). Although more research is still needed to understand the exact mechanisms by which bilingual instruction may promote English acquisition (see Umansky & Reardon, 2014 for possible explanations that may include socio-emotional, self-esteem, community-based, and identity-construction benefits), existing studies have provided compelling evidence that EL children are capable of learning English in addition to their L1 (National Academies of Science, Engineering, and Medicine, 2017). Given these findings and the widely reported cognitive and behavioral benefits of bilingualism (Bialystok, 2001; 2011), policymakers and educators in Michigan may want to consider incentivizing the development or expansion of bilingual programs, particularly dual-bilingual programs (Robinson-Cimpian et al., 2016), and particularly within districts that have concentrations of children with shared home languages. With the rapid expansion of online learning across the curriculum due to COVID-19, such programs could be envisioned as hybridized or online in the future, and potentially shared across districts or schools (without borders, so that even physically dispersed, shared-home-language learners could participate in at least some aspects of bilingual education.

### **10.3 Carefully Investigating and Monitoring the Effect of retention**

In 2016, Michigan passed a third-grade reading retention law that stipulates that starting from 2020, EL and non-EL third graders who test below grade level on Michigan's reading assessment (i.e., the Michigan Student Test of Educational Progress – English Language Arts) shall be slated to repeat third grade. There is much controversy surrounding this law, particularly with regard to its impact on EL children (see Winke & Zhang, 2019). Such controversy contributes to a long-standing and ongoing debate as to whether retention is an effective remedial measure for low-achieving students (Schwerdt, West, & Winters, 2017; other researchers have also discussed the psychological effects of retention on children, e.g., Anderson, Jimerson, & Whipple, 2005).

This study investigated the effect of traditional retention on ELs' time to proficiency using longitudinal data collected prior to the implementation of Michigan's reading retention law. Consistent with many previous studies, findings of this study suggest that retention is associated with more negative than positive impacts on ELs in the long term (Jimerson & Renshaw, 2012). These findings serve as a timely warning to Michigan educators and policymakers who are enthusiastic about test-based retention. Despite their overt optimistic attitudes toward retention, this research provides little support for the idea that retention would effectively address student underachievement (Jimerson & Renshaw, 2012). Proponents of retention may point out that more recent studies examining test-based retention have found overwhelming positive effects on EL and non-EL students (e.g., Schwerdt et al., 2017; Figlio & Özek, 2020). However, one should be aware that these positive effects may not be attributable to retention per se, given that researchers of these studies were not able to disentangle the effect of retention from the effect of additional instruction services that retained students were able to receive. Therefore, rather than

pinning their hopes on retention, Michigan educators and policymakers should carefully consider how they can improve the quality of education for students who are at risk for academic failure to effectively help those students. In addition, they should use this study's results as a baseline to carefully evaluate and monitor the effect of test-based retention on ELs (e.g., whether test-based retention triggers more incidence of retention among ELs, whether test-based retention facilitates or hinders EL children's English acquisition as compared with traditional retention).

#### **10.4 Strengthening Writing Instruction for ELs**

Analyses of individual proficiency criteria showed that writing represented the largest barrier to proficiency for EL children in Michigan public schools. This is different from previous research that found reading to be the most difficult skill for ELs (e.g., Thompson, 2017). It may also differ from the expectations of most EL teachers and educators. However, this finding is not completely surprising considering that writing has received much less attention than reading from EL researchers and educators (Gillanders, Franco, Seidel, Castro, and Mendez, 2017; Olson, Scarcella, & Matuchnik, 2015). For example, in a comprehensive review of studies on the language and literacy of dual language learners, Hammer et al. (2014) found that among a total of 182 articles reviewed, only 6 were conducted on the writing development of dual language learners. The small body of research on EL writing suggests that the writing development of EL children is roughly similar in progression to that of monolingual speakers (William & Lowrance-Faulhaber, 2018). Like their monolingual English-speaking peers, ELs may encounter a variety of challenges in learning to write and would need high-quality instruction on aspects such as word structure and meaning, syntactic knowledge, spelling (Perin, De La Paz, Piantedosi, & Peercy, 2017). In addition, ELs differ from their monolingual peers in that their composing processes are uniquely affected by their L1 language and literacy skills (William & Lowrance-

Faulhaber, 2018). Research has suggested that underdeveloped L1 skills may negatively affect the English writing acquisition of ELs, who may need to draw on their L1 knowledge during the early stage of English writing development (Lanauze & Snow, 1989; Savage, Kozakewich, Genesee, Erdos, & Haigh, 2017). Findings of this study should serve as an alarm call for EL educators and teachers to strength writing instruction for ELs. To do so, teachers may need to provide opportunities for ELs to develop their L1 proficiency at school. In addition, test developers should pay more attention to how writing is taught and practiced in the classroom so that they can develop writing assessments that are more closely aligned with the curriculum and classroom writing instruction. This will contribute to a seamless move from learning to assessment, which helps improve test validity and washback (Messick, 1996).

Chapter summary:

- School accountability systems should be improved to reflect the diversity of the EL subgroup.
- Policymakers, educators, and parents should take more measures to support ELs' L1 development.
- The effect of retention should be carefully investigated and monitored.
- Writing instruction should be strengthened for ELs.

## CHAPTER 11 LIMITATIONS

This study has at least five limitations. First, results of this study are descriptive and cannot support causal inferences. This limitation should be kept in mind particularly when interpreting the effects of instructional programming and retention on time to proficiency. Although I controlled for important student and school covariates (e.g., cohort, initial proficiency, school fixed effects) that may have affected instructional program assignment, it does not necessarily eliminate the selection bias, and thus the results should not be interpreted as causal assessments of the effectiveness of different programs on ELs' English acquisition. Similarly, the effect of retention on students' likelihood of attaining proficiency cannot be interpreted as causal because the exact reasons why students were retained in grade were not recorded or provided by the Michigan Education Research Institute (MERI). Future research should employ a randomized experiment or rigorous quasi-experimental design to understand the causal relationships between time to proficiency and factors that affect this time (see Slavin et al., 2011 for a discussion of how a randomized experiment should be designed to investigate program effectiveness).

Second, the proficiency standards for ACCESS for ELLs (i.e., Michigan's state English language proficiency assessment, hereafter ACCESS) and the proficiency criteria that the Michigan Department of Education (MDE) established for ACCESS (composite score and individual domains) changed over the course of the study. These changes could to some extent limit the generalizability of this study because the criteria used by the MDE over the study period may not represent the criteria they use in other years or the criteria used by other WIDA states during the same time period (see Linquanti & Cook, 2015). Within the specific context of survival analysis, adjusting proficiency criteria can have a direct impact on the estimate of time to proficiency/reclassification: whereas including additional proficiency criteria and/or adjusting



the cut scores to a higher level (while holding other things constant) would likely result in a longer estimate of time to proficiency, reducing the number of proficiency criteria and/or adjust the cut scores to a lower level may lead to the opposite result. For example, the MDE re-adjusted the proficiency criteria for ACCESS in 2020 by dropping all proficiency criteria for the four domains and only keeping the criterion for the composite score. Given that writing was a larger barrier than the composite score for Michigan ELs in early elementary grades, such a decision may lead to shorter times to proficiency/reclassification for ELs who are enrolled in Michigan public schools in or after the year of 2020 than for those students observed in this study. Although adjustment to proficiency criteria adds complexity to investigations of ELs' longitudinal outcomes, readers should keep in mind that the MDE makes changes to the proficiency criteria for ACCESS every year in order to improve the validity of the interpretations and uses of the test scores. Thus, the changes should have benefits that far outweigh, perhaps, the researcher's needs and interests in an as-uninterrupted-as-possible longitudinal dataset. Longitudinal datasets are interrupted or broken in the real world often and for many various and valid reasons, and thus statistical processes exist to cope, as best possible, with such realities. In addition, readers should note that the six-level system of WIDA's English Language Development Standards kept its meaning in terms what EL children can do at each level throughout the entire study, which is one way in which the interruption can be seen as having minimal impact.

Third, the English placement test scores that ELs obtained upon school entry were not available from the MDE. Such scores would have been a more ideal Year 0 indicator. Previous studies suggest that ELs' initial proficiency levels in English are significantly related to their later development in English language and literacy skills (Kieffer, 2008; Kieffer & Parker, 2016;

Slama et al., 2017; Thompson, 2017). Meanwhile, initial English proficiency may be confounded with EL program placement in a systematic manner. Although I controlled for ELs' initial proficiency by using their end-of-kindergarten ACCESS scores as a proxy for Year 0 scores, this is not an ideal solution. An important implication for the MDE is that they should better track ELs' progress toward English proficiency by improving their data collection and management systems. Failing to collect or store ELs' English placement test scores limits educators, researchers, and policymakers' abilities to understand the variation in ELs' initial proficiency and how such variation is associated with ELs' later English development. Researchers have long recommended that policymakers take ELs' initial proficiency levels into account in establishing the expected time frames for attaining proficiency in English and content areas (Cook, Linquanti, Chinen, & Jung, 2012; Thompson, 2017). Implementing this recommendation requires that states systematically collect and store the initial proficiency data for each EL (see Slama et al., 2017 for how Texas has implemented this idea in its EL policies).

Fourth, this study was not able to track ELs' progress toward proficiency beyond elementary school. This is because Michigan joined the WIDA consortium relatively recently (in 2013) and has not accumulated enough proficiency data to allow for a retrospective data collection over a longer time span. Studies examining time to reclassification indicate that reclassification slows in middle school (Umansky & Reardon, 2014). One would expect to see a similar trend in the likelihood of proficiency attainment, given that students with the highest levels of English proficiency are most likely to be reclassified as former ELs during elementary school. With the influx of new data in the coming years, researchers should investigate how the likelihood of proficiency changes for ELs in Michigan public schools from the elementary through high school years, and if possible, going beyond K-12 to track ELs' achievement in the

adulthood. This will help educators and researchers understand the trajectories of English development for ELs with disabilities and long-term ELs (i.e., ELs who do not successfully attain English proficiency after having spent six or more years in U.S. schools). This line of research will also help determine early antecedents that predict later success/non-success in ELs to understand the real-world meaning of these data. A significant interruption to this work, however, will be COVID-19, which presented a generation of children with missing test scores in at least 2020 and 2021, learning delays or growth aberrations unrelated to traditional educational and home factors, and schisms in academic programming such that educational categorizations may need to be tagged as separate and unique within some later-to-be-defined, COVID-19 time frame. More advanced statistical methods such as discontinuous piecewise latent growth modeling may need to be used to analyze data that spanning pre- and post-COVID periods.

Finally, although this study was able to keep track of EL children who switched to different public schools within Michigan over time, it was not able to track ELs who left the Michigan public school system. The results therefore may not generalize to students who moved to other states or transferred to private schools or home schooling before they attained proficiency.

## CHAPTER 12 CONCLUSION

For this dissertation project, I estimated the length of time it took English-learner (EL) children in Michigan public schools to attain proficiency and investigated how six factors of interest (i.e., primary disability type, primary home language, instructional programming, poverty status, home English use, and retention) were related to variation in ELs' time to proficiency. I did so by analyzing the demographic and standardized test data from six cohorts of students who entered Michigan public schools as ELs in kindergarten between 2013-2014 and 2018-2019. Unlike previous studies, I situated this study in a state that has been a member of a large English-language-proficiency (ELP) assessment consortium (i.e., WIDA). In addition, I used a novel approach and investigated time to proficiency independently from reclassification criteria that have been debated as irrelevant to the core construct of ELP. These improvements allow for wider extrapolation of the findings. The data indicated that half of students entering Michigan public school as ELs in kindergarten took around five years to attain proficiency in English. Findings of this study further indicated ELs' time to proficiency was significantly related to their sociodemographic and educational backgrounds. Specifically, ELs with disabilities, ELs speaking some particular home languages (e.g., Arabic, AfroAisatic, and Spanish), and ELs who ever lived in poverty were less likely to attain proficiency than their peers who did not have those backgrounds. In addition, ELs who received partial instruction in their home language (L1) and ELs who never used English at home were equally likely, and in some cases slightly more likely, to attain proficiency as compared with their peers who were immersed full-time in an English-only environment at school and at home. Lastly, students who were ever retained in grade were less likely to attain proficiency than their peers who were never

retained in upper elementary grades despite short-term gains associated with retention in early elementary grades.

This research is important because it contributes to a small body of literature on child second language acquisition (SLA) by promoting the understanding of how EL children develop English language proficiency over time and how heterogeneity within this population is related to differences in their proficiency development. Considering that current education policies (e.g., ESSA) are providing states with more flexibility in establishing their own accountability systems, results of this study can be used by Michigan and perhaps other WIDA states to establish more appropriate English-proficiency targets for ELs from diverse backgrounds. As states do so, it is important that they conduct ongoing examination of longitudinal EL data to carefully monitor and evaluate the effect of their policies on ELs and schools that serve them. In addition, future researchers should pay more attention to child SLA and factors that influence child SLA. It is important that researchers investigate factors at multiple levels, including child-internal factors, and contextual factors at the micro-, meso, and macro-levels (Umansky, Callahan, & Lee, 2020) that may directly or indirectly affect a child's language learning trajectory. One particular gap in the current child SLA literature is the longitudinal (as opposed to cross-sectional) relationship between first language (L1) development and second language (L2) development. A better understanding of this issue will help develop more robust theories of child SLA. It will also help inform important policy aspects of EL education, including L1 maintenance and bilingual education.

## **APPENDICES**

## APPENDIX A Data Preparation

### A.1 The Michigan Department of Education (or MDE)'s Test Data Cleaning Process

The MDE's test dataset consisted of 179,021 observations for 57,327 students in a person-period (or long) format. Table 30 lists all MDE variables that I used for this study and the number of missing records on each of these variables.

Table 30. A list of study variables and their origins in the MDE data

Study variable	MDE variable	Missing
School year	SCHOOL_YEAR	0
Grade	TESTED_GRADE	0
Reading scale score	RDG_SS	1,167
Reading proficiency level	RDG_PL	1,167
Writing scale score	WRI_SS	1,677
Writing proficiency level	WRI_PL	1,677
Listening scale score	LISTENING_SS	1,137
Listening proficiency score	LISTENING_PL	1,137
Speaking scale score	SPEAKING_SS	1,580
Speaking proficiency level	SPEAKING_PL	1,580
Overall composite scale score	OVERALL_SS	2,058
Overall composite proficiency level	OVERALL_PL	2,058
School code	TESTED_SCHOOL_CODE	0
District code	TESTED_DISTRICT_CODE	0

Although ELs are supposed to take ACCESS annually, I found that 32 students took ACCESS twice in the same year. The MDE provided the following explanation for repeated ACCESS administrations within the same year:

We often have it happen that students switch between districts during the 7-week WIDA ACCESS testing window. Sometimes that accounts for duplicate testing records. At times it can happen within the same district, but that's more unlikely. And there are times when educators don't know that a student should be taking WIDA ACCESS and not WIDA Alternate ACCESS (or vice versa) and when they discover their error they make the correction which results in duplicate records.

MERI (Email communication with Jasmina Camo-Biogradlija, 2020)  
For the purpose of this study, I retained the record with a higher overall composite score for each of the 32 students for those years they had two sets of ACCESS results. Descriptive statistics revealed 872 Alternate ACCESS test records (produced by 434 students). I code these as missing because they could not be evaluated against the MDE's proficiency criteria for ACCESS (note that the MDE established proficiency criteria for Alternate ACCESS in fall 2020). I also corrected the grade information for two cases where human data-entry errors had likely occurred

(i.e., the grade level proceeded backward over time). The final MDE dataset contained only observations with complete test data, totaling 175,768 records from 56,039 students.

## A.2 CEPI's Demographic Data Cleaning Process

The CEPI data file contained 214,412 records from 59,615 students, organized in a person-period (or long) format. Table 31 lists all CEPI variables that I used for this study: five are time-invariant variables, and six are time-varying variables. Students typically had one record for each year (or period) they were observed; however, for students who transferred between schools in a particular year, they had as many records for that year as the total number of schools they were enrolled in, one record for each school. The dataset records time-varying school information because CEPI collects demographic data at individual schools three times every year, once in the fall, once in the spring, and once at the end of the academic year. Values of other variables may also vary within a particular year and across the years. For example, a student may self-report being male at one data-collection occasion and report being female at another data-collection occasion.

Table 31. A list of study variables and their origins in the CEPI data

Study variable	CEPI variable	Missing records
(time-invariant)		
Gender	STABLE_GENDER	0
Race/ethnicity	STABLE_RACE_ETHNIC	0
Age	STABLE_DATA_BIRTH	0
Primary home language	PRIMARY_LEP_LANGUAGE_SY	13,598
Primary disability type	PRIMARY_DISABILITY_CODE_SY	193,715 <sup>a</sup>
(time-varying)		
Grade	GRADE_SY	0
Poverty status	IS_ECONOMICALLY_DISADVANTAGED	0
English use at home	LEP_LANGUAGE_CODE_1-5	17,364
Program	LEP_INSTR_PROG_CODE_1	173,64
School code	SCHOOL_CODE	0
District code	OPERATIONAL_DISTRICT_CODE	0

*Note.* <sup>a</sup>Missing values were assumed to indicate no disabilities.

For convenience of data analysis, my goal was to clean the CEPI data to obtain a dataset (a) where each student only had only one record for each school year they were observed, and (b) where only time-varying variables had time-varying values. One key step was to cope with within-person data variation. CEPI had already handled most within-year, within-person data inconsistencies (with the exception of school-transfer-related data changes); however, I still needed to deal with cross-year, within-person data variation. For time-varying variables, this is not an issue because cross-year, within-person data variation is expected. However, for time-invariant variables like gender, this poses a problem unless one wants to treat those variables as



time-varying. Table 32 displays a fake case that has inconsistent within-person information on time-invariant variables (with inconsistent information in bold).

Table 32. A fake case with inconsistent cross-year, within-person information on time-invariant variables

ID	School Year	Grade	Gender	Age	Primary disability type	Primary home language	Race/ethnicity
100001	2013-2014	0	Female	5.5	2	Arabic	White
100001	2014-2015	1	<b>Male</b>	5.5	<b>1</b>	Arabic	White
100001	2015-2016	2	Female	5.5	2	<b>Spanish</b>	White
100001	2016-2017	3	Female	5.5	2	Arabic	White
100001	2017-2018	4	Female	<b>5.0</b>	2	Arabic	<b>Hispanic</b>
100001	2018-2019	5	Female	5.5	2	Arabic	White

Inconsistent within-person information may have occurred due to human data-entry error or true changes in student status. Because I was not able to decide if the inconsistency arose from human error (since no other official sources of demographic information were available), I chose to believe that all inconsistent instances represented true changes in the data. However, this does not mean that variables like gender, age, and race/ethnicity are naturally time-varying. For instance, a person's gender does not naturally change over time like their height or weight, although gender may appear to be time-varying in a large dataset due to a small number of special cases who have changed their gender. Because I am not interested in the effects of time-varying gender, age, primary disability, primary home language, or race/ethnicity, I decided to treat those variables as time-invariant by resolving within-person inconsistencies on those variables. Specifically, I unified inconsistent data for each time-invariant variable by using its most frequent within-person value. For example, for the fake case in Table 32, I would determine that the gender is female, age is 5.5, disability type is 2, primary home language is Arabic, and race/ethnicity is White because those values are more frequent. In very few cases where there was a tie between two or more values, *R* was programmed to select one value randomly. Table 33 summarizes the number of cases with inconsistent cross-year data for each time-invariant variable listed in Table 31. As Table 33 shows, very few students had inconsistent demographic information on time-invariant variables. This suggests good data quality and that whether treating those variables as time-invariant or time-varying would not have much effect on the results.

Table 33. Numbers of cases with inconsistent data on time-invariant variables

Time-invariant variable	Number of cases with inconsistent cross-year data	Number of records affected
Gender	9 (0.015%) out of 59,615	48 (0.022%) out of 214,412
Race/ethnicity	132 (0.214%) out of 59,615	895 (0.417%) out of 214,412
Age	22 (0.037%) out of 59,615	139 (0.065%) out of 214,412
Primary home language	1372 (2.302%) out of 59,599 <sup>a</sup>	7007 (3.268%) out of 214,380 <sup>a</sup>
Primary disability type	789 (1.323%) out of 59,615 <sup>b</sup>	4078 (1.902%) out of 214,412 <sup>b</sup>

*Note.* <sup>a</sup>I fixed data inconsistencies for primary home language after I coded this variable based on the coding scheme described in the method section. Note that 16 students (32 records) did not have any data for primary home language. For students with home language data for some but not all the years, I mutated the missing records ( $n = 13,566$ ) for those students by using the most frequently reported value (e.g., if a student has four home language records, one being Chinese, two being Spanish, and one missing, the student would be coded as a Spanish speaker for all four records).

<sup>b</sup>I fixed data inconsistencies for primary disability type after I coded this variable based on the coding scheme described in the method section. Note that 7,125 students (28,695 records) were identified with at least one type of primary disability. The remaining students ( $n = 52,490$ ) had no value for primary disability type and were assumed to have no disabilities.

To create a dataset that had one record per student per each school year, I also had to deal with within-year, within-person variation in school code. Specifically, I used the variable `DAYS_ATTENDED` (which indicates the number of days a student was enrolled in a given school in a particular year) to determine the primary school for students who switched between schools within the same academic year. The primary school was defined as the one that the student attended for most days within a particular year. This led to the removal of 13,596 records, leaving 200,816 in the dataset (produced by 59,615 students).

A final note about the CEPI data is that reclassification status cannot be reliably drawn from the dataset. There are two time-varying variables indicating a student's EL status, `LEP_EXIT_DATE` and `LEP_REENTRY_DATE`. However, I found that some students who were exited from LEP in a year without re-entering LEP were still administered ACCESS in subsequent years. The MDE explained that this may have happened for two reasons,

- (a) Scenario 1: Student was exited at the beginning of the school year but took the WIDA ACCESS that same school year anyway. Our Accountability measures related to the federal requirement that ALL ELs take the summative ELP assessment are based on student identifications as EL at ANY point in that MSDS school year. So, even if a student exited from services in the fall, they were technically still EL DURING that school year. Therefore, we would have expected them to take the test.
- (b) Scenario 2: Student was exited at the beginning of the school year but continued to take the ACCESS in subsequent school years and they do not have a LEP Reentry Date (excluding the current school year in which they were exited). This is more likely a

scenario of the right hand not knowing what the left is doing in the district or a technology error. This could be an error that occurred in their student information system that, at the local level, reidentified the student as EL so therefore they continued to test the student. It could be that the district EL coordinator knows the student was exited but that wasn't communicated to the building level coordinator so they think the student is EL and continue testing the student. In any of these cases though, it's possible that the student doesn't actually have a valid score. I would look to see in your data if the scores are set to valid or not. In cases where a student took the test but was not who we would consider a formally identified EL we would invalidate the test scores.

MERI (Email communication with Jasmina Camo-Biogradlija, 2020)

In addition, some students who had met all proficiency requirements for ACCESS in a given year and were no longer administered ACCESS in subsequent years (presumably because they were reclassified as former EL) had no LEP\_EXIT\_DATE. The fact that those students remained in the CEPI dataset indicated that they did not leave Michigan public schools. This may have happened because the local educational agency failed to put down the LEP\_EXIT\_DATE for those students. (However, this has not been confirmed by the MDE.) Lastly, reclassification usually occurs in summer. I believe the dataset I requested in September 2019 did not contain the LEP\_EXIT\_DATE for most students who were reclassified at the end of the 2018-2019 academic year, the last year of the study period. (However, this has not been confirmed by the MDE.)

### A.3 Merging the MDE and CEPI datasets

I merged the cleaned MDE and CEPI datasets based on a unique student identifier and school year. The merged dataset (Merged\_1) had 200,817 observations from 59,615 students in a long format. There are three overlapping variables between the MDE and CEPI datasets: grade, school code, and district code. Among the 200,817 observations, 39 had different TESTED\_GRADE (MDE) and GRADE\_SY (CEPI), 890 had different TESTED\_SCHOOL\_CODE (MDE) and SCHOOL\_CODE (CEPI), and 1951 had different TESTED\_DISTRICT\_CODE (MDE) and DISTRICT\_CODE (CEPI). I went with the TESTED\_GRADE data unless there was clear evidence for data-entry error (e.g., grade proceeded backwards). Because the school and district in which students took ACCESS in a particular year were not necessarily their primary school or district in that year, I used the school and district information from CEPI whenever there was a conflict between the two datasets.

The fixed grade information was used to determine student cohort. There were in total six cohorts, defined by the year the student entered kindergarten. For example, the 2013-2014 kindergartener cohort included all students who entered kindergarten in 2013-2014. Two students (eight observations) were removed from the merged dataset because their grade information indicated they entered kindergarten prior to the first year of the study period (i.e., 2013-2014). In other words, their times of entry into Michigan public schools were not observed. The resulting dataset contained 200,808 records from 59,613 students (Merged\_2).

I first searched for any student who entered the study late, that is, who did not have test data until one or more years after they entered kindergarten. I found that 4809 students (13,685 observations) did not have any test data for the first year at school (i.e., kindergarten). Among those 4809 students, 2576 had no data at all, 2005 had missing data only for the kindergarten

year, and 228 had missing data for the kindergarten year and one or more of the other years while enrolled in school. The reason behind the missing data is unknown. Although ELs are allowed to be exempt from content-area assessments within the first three years at school, they are not supposed to be exempted from English Language proficiency (ELP) assessments. One possibility is that those students were identified with severe disabilities that prevented them from taking the regular ACCESS (at least in the first year) so they took other ELP assessments (e.g., ACCESS Alternate) instead; or they did take ACCESS but did not receive valid scores. Because what happened to those students in their kindergarten year was unknown and could have influenced their English acquisition in subsequent years, those kids were not considered in subsequent analyses.

The dataset (Merged\_3) following the removal of those 4809 students had 187123 records for 54804 students. Within this dataset, I looked for students who left the study early, either temporarily or permanently, prior to the end of the data-collection period (i.e., 2018-2019). Some reasons that may account for students leaving the study (or the merged dataset) early include (a) reclassification (students attained ELP and all other reclassification criteria and became reclassified as proficient in English), (b) leaving Michigan public schools (students switched to private schools or home schooling, moved out of state, or dropped out of school), (c) having missing test data (students obtained invalid ACCESS scores, lost some scores due to human error, or took ACCESS Alternate or other ELP assessments instead of ACCESS).

I first removed 17,325 records that did not have any test data. I then identified 1,496 students who left the merged dataset temporarily. Following Kieffer and Parker (2016), I kept those students' records up to the year right before they left Michigan public schools (or had missing data) for the consideration that their data after returning to the dataset may have differed qualitatively from the data of their counterparts who had never left the dataset (in terms of schooling, testing, or other ELP-related experiences). This led to the removal of 2,658 observations. Then I set out to find students who left the dataset permanently prior to the last data collection year, 2018-2019. In total, I found 11,843 students (31,688 observations) who left the dataset early permanently, with 5,437 leaving one year before the data collection ended, and 6,406 leaving more than one year prior to the end of data collection. Because this study stopped following students after they left the dataset, its results cannot be generalized to those students who attained proficiency after they transferred to alternative school systems, moved out of Michigan, dropped out, or had missing test data.

The above cleaning process led to a dataset that consisted of 167,140 records from 54,804 students (Merged\_4). Then, I checked the amount of missing data on each demographic variable of interest. All variables except *program* and *English use at home* had complete data. I thus removed 2,687 records from 658 students who had *program/English use at home* data missing for some or all periods. In other words, only cases with complete data on all variables were retained for subsequent coding and analyses.

The resulting dataset (Merged\_5) contained 164,453 observations from 54,146 students. I identified in this dataset those students who attained proficiency as measured by ACCESS according to the MDE's proficiency criteria as follows:

- 2013-2014 through 2014-2015: A minimum overall composite score of 5.0 (out of 6.0) on ACCESS, and a minimum score of 5.0 (out of 6.0) in the four domains of speaking, listening, reading, and writing
- 2015-2016: A minimum overall composite score of 5.0 (out of 6.0), and a minimum score of 4.5 (out of 6.0) in the four domains of speaking, listening, reading, and writing

- 2016-17 through 2018-2019: A minimum overall composite score of 4.5 (out of 6.0), and a minimum score of 4.0 (out of 6.0) in reading and writing

Based on the criteria above, 12,430 students attained ELP during the study period. The remaining 41,716 students were censored because their event time (i.e., the time when they attained ELP) was undetermined. Censoring is a critical concept in survival analysis. In this study, a student is considered censored if the student

- (a) had not attained proficiency by the end of the study period (the student either never attains proficiency or attains proficiency after 2018-2019, when data were not available)
- (b) had not attained proficiency before they left the study (e.g., switched to private schools or home schooling, dropped out, moved out of state, took other ELP assessments)
  - 1. had not attained proficiency prior to leaving temporarily
  - 2. had not attained proficiency prior to leaving permanently

Rather than deleting censored students, survival analysis allows those students to remain in the risk set (the group of students who have not attained proficiency but are eligible to do so) and to contribute to the estimated risk of ELP (or other outcome) attainment in each time period they were observed. This helps avoid bias in estimates of time to ELP attainment. The number of censored students is presented below by the mechanism of censoring.

- Mechanism (a), censored due to the end of data collection: 34,621 (82.99%)
- Mechanism (b1), censored due to leaving the study early temporarily: 1,404 (3.37%)
  - 745 left after one year
  - 392 left after two years
  - 216 left after three years
  - 51 left after four years
- Mechanism (b2), censored due to leaving the study early permanently (prior to the end of data collection): 5,691 (13.64%)
  - 2,885 left after one year
  - 1,446 left after two years
  - 810 left after three years
  - 415 left after four years
  - 135 left after five years

Table 34 displays the number of students who attained ELP and the number of students who were censored by the number of years since kindergarten entry. To prepare this dataset for survival analysis, I coded all records post to first ELP attainment as missing and removed those records. The final dataset (Merged\_final) had 161,211 records from 54,146 students.

Table 34. The number of students who were censored and the number of students who attained ELP by the number of years since the student entered kindergarten

Years since school entry	Time interval	Starting number	Event = 1 (attain ELP)	Censored due to end of data collection	Censored due to other reasons	Total censored at the end of the year	Event proportion	Survival rate
0	[0, 1)	54146	NA	NA	NA	NA	NA	1
1	[1, 2)	54146	2062	7950	3630	11580	3.81%	96.19%
2	[2, 3)	40504	686	7548	1838	9386	1.69%	94.56%
3	[3, 4)	30432	1724	7034	1026	8060	5.67%	89.21%
4	[4, 5)	20648	2898	5792	466	6258	14.04%	76.69%
5	[5, 6)	11492	3776	3592	135	3727	32.86%	51.49%
6	[6, 7)	3989	1284	2705	0	2705	32.19%	34.92%
<b>Total</b>			<b>12430</b>	<b>34621</b>	<b>7095</b>	<b>41716</b>		

The entire data cleaning process is illustrated in Figure 17.

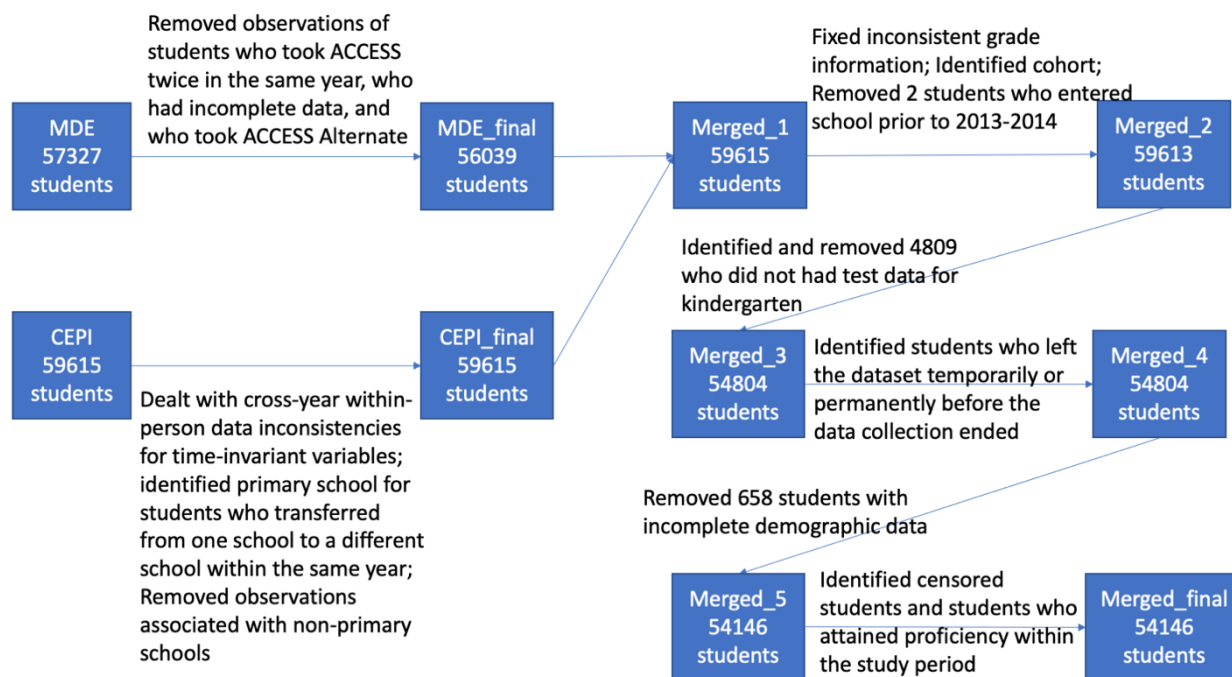


Figure 17. Visualization of the data cleaning and preparation process

## APPENDIX B Considerations in Coding Predictors

In this section, I describe the considerations I took in coding the following predictors of interest: primary disability type, primary home language, instructional program, English use at home, poverty status, and retention.

### B.1 Primary Disability Type

In accordance with the requirement of the Individuals with Disabilities Education Act (IDEA), the Michigan Department of Education (MDE) identified all students, including ELs, with the 13 categories of disabilities (shown in Table 35) during the period of data collection.

Table 35. Disability categories and the numbers of records and students in each category

Number	Disability category	Number of records	Number of students
1	Cognitive Impairment	1135	385
2	Emotional Impairment	260	91
3	Hearing Impairment	394	128
4	Visual Impairment	92	29
5	Physical Impairment	183	67
6	Speech & Language Impairment	11611	4464
7	Early Childhood Developmental Delay	1198	673
8	Specific Learning Disability	2904	1286
9	Severe Multiple Impairment	139	38
10	Autism Spectrum Disorder	1689	591
11	Traumatic Brain Injury	15	5
12	Deaf-Blindness	1	1
13	Other Health Impairment	1076	428

For the purpose of this study, I treated primary disability type as time-invariant despite the fact that some ELs' primary disability type did change over the years of data collection (789 out of 7125 disabled ELs). As I described in the Appendix A *Data Preparation*, for an EL who was identified with varying primary disabilities over the years, I decided that the most frequent type of disability was the EL's primary disability, treated as time-invariant (a random type was selected when there was a tie between two or more types). I made this decision for two considerations. First, existing literature suggests that the disability identification process is error prone for ELs (Sanatullova-Allison & Robison-Young, 2016) due to difficulties in differentiating between ELs with learning disabilities and those with English acquisition issues (Kangas, 2019; Klingner, Artiles, & Barletta, 2006). Given the literature's description of the "serious and pervasive problem" of misidentification of ELs for special education (Sanatullova-Allison & Robison-Young, 2016, abstract), it is possible that an EL with time-varying disabilities was *misidentified* with a less-frequently-reported disability that the student did not have in addition to a primary disability that the student was most frequently identified with over the years. If this is

the case, then the most frequent disability would likely best represent the EL's disability profile. However, I acknowledge that not all within-student time-varying disabilities suggest misidentification; some ELs with varying disabilities may have truly suffered from different disabilities during different time periods. However, because it is hard to determine the actual disability pattern for each individual with time-varying disabilities, I decided to treat primary disability type as time-invariant rather than time-varying considering that this approach reduces the level of complexity in data analysis and interpretation. Analytical convenience was therefore the second consideration weighed on this decision.

Following Kieffer and Parker (2016), I only focused on two types of primary disabilities in this study: specific learning disability and speech and language impairment. This is because these two types of primary disabilities are most commonly found among EL students and need to be diagnosed using language assessments. All other primary disability types were coded "other."

## B.2 Primary Home Language

The sampled students represented 245 identified primary home language groups. Because it is difficult to handle this large number of home language subgroups in statistical analyses, I came up with two methods to recode primary home language backgrounds. The first method was informed by previous studies (e.g., Thompson, 2017), whereby I identified the 10 most frequently reported home languages and coded the remaining home languages as "other." The second method was based on the concept of linguistic distance (Chiswick & Miller, 2005). Using this method, I collapsed the languages that are linguistically closer to each other based on the language family information available from *glottolog* (<https://glottolog.org>). I did so to reduce the large number of home language groups into a small, manageable number of home language family groups.

The first coding method led to the following 11 home language categories: Spanish ( $n = 19,591$ ), Arabic ( $n = 14,061$ ), Bengali ( $n = 1,539$ ), Chinese ( $n = 1,510$ ), Albanian ( $n = 1,060$ ), Aramaic ( $n = 1,025$ ), Japanese ( $n = 1,030$ ), Telugu ( $n = 1,255$ ), Urdu ( $n = 860$ ), Vietnamese ( $n = 971$ ), and "other" ( $n = 11,244$ ). As one may have observed, the category of "other" is quite sizable and non-informative.

The second coding method partially solves this problem. For this method, I retained the coding for the five largest home language groups (i.e., Spanish, Arabic, Chinese, Bengali, Albanian), and I recoded the remaining home language groups into a small number of home language family groups. Specifically, I first identified all home languages (except for the top five;  $n = 59$ ) that had 100 or more data records (roughly spoken by 30 students); then, I classified those languages into language families according to *glottolog*. Top-level language families ( $n = 244$ ) were used to do the coding; however, because the Indo-European language family was too large (with over 30,000 records) to be informative, level-two language families were used instead to code all Indo-European languages. This resulted in 21 home language families, including 14 top-level and seven level-two families. I then identified the five largest home language families (all with more than 4000 records). I also created two additional home language families of similar size by collapsing the Italic and Germanic language families into the Italic/Germanic language category, and the Japonic and Koreanic families into the Japanese/Korean category. In total, I identified seven large language families among the less frequently reported home language groups:

- 3) AfroAsiatic: Amharic, Aramaic, Oromo, Somali, Syriac, Tigrinya



- 4) Austronesian: Filipino, Indonesian, Central Khmer, Malay, Philippine Languages, Tagalog, Vietnamese
- 5) Dravidian: Kannada, Malayalam, Tamil, Telugu
- 6) BaltoSlavic: Bosnian, Bulgarian, Macedonian, Polish, Russian, Serbian, Ukrainian
- 7) IndoIranian: Gujarati, Hindi, Kurdish, Marathi, Nepali, Panjabi, Persian, Pushto, Urdu
- 8) Italic/Germanic: French, Italian, Portuguese, Romanian, English
- 9) Japanese/Korean: Japanese, Korean

All other languages that do not belong to the seven language families (or the five most frequently reported languages) were coded “other.” Although this method still evoked the category of “other,” the size of this category ( $n = 3,128$ ) was substantially reduced compared to the first method.

To test the statistical efficiency of the two coding methods, I compared two alternative discrete-time hazard models that predicted the outcome of English proficiency attainment using time (categorical variable) and primary home language: one model used method 1 to code the primary home language variable (Model LangA), and the other model used method 2 to do so (Model LangB). Because the two models are not nested, I used the Bayesian information criterion (BIC) to compare their goodness-of-fit to the data (Singler & Willet, 2003, p. 401). BIC is a commonly used model-selection criterion derived from information theory. When selecting from among several candidate models, one would want to prefer the model that best approximates the data, or the model with the lowest BIC (optimal model). Other candidate models in the pool can be evaluated against the optimal model using the delta BIC, which corresponds to the difference in BIC between the best model and a particular candidate model in the pool. The following guidelines are usually applied to assess the evidence against a candidate model (with a higher BIC) being the best model (Kass & Raftery, 1995): If the delta BIC is

- less than 6, the evidence against the candidate model with a higher BIC is positive
- between 6 and 10, the evidence is strong
- greater than 10, the evidence is decisive

One may also use the Akaike Information Criterion (AIC) for the same purpose; however, I chose to use BIC because BIC not only penalizes the loglikelihood statistic for the number of parameters in the model as AIC does, but it also takes into account total sample size (Singler & Willet, 2003, p. 402). The results showed that the BIC of Model LangB was 750 points smaller than that of Model LangA, providing strong evidence that Model LangB was a better-fitting model.

### B.3 Instructional Programming

In this study, I investigated the effect of instructional programming on ELs’ time to proficiency following the coding method used by Umansky and Reardon (2014). Specifically, I coded program as a time-invariant variable representing the EL’s initial program placement in kindergarten in one of the four types of programs: ESL program, transitional bilingual (TB) program, dual-language bilingual (DB) program, and no program (NP). Although this method did not use all the information that I had about what program each student was in each year, the results did not differ much from when I coded program as a time-varying variable. This is perhaps because the overwhelming majority of students remained in their initial EL program until that program ended by design or they left the dataset (see Table 4 in the Chapter 7 for the 13 most frequently observed program patterns in the 2013-2014 kindergarten cohort).

In addition, I found that coding program as a time-invariant variable was associated with higher statistical efficiency than coding it as a time-varying variable. Specifically, I compared the goodness-of-fit of two models that used time (a categorical variable), program, and the interaction between time (a continuous variable representing the number of years that the student has remained in Michigan public schools) and program to predict the outcome of English proficiency attainment. In the first model (Model ProgramTV), program was coded as a time-varying variable indicating the type of program the EL was enrolled in a particular time period. In the second model (Model ProgramTI), program was coded as a time-invariant variable indicating the type of program the EL was initially placed in upon kindergarten entry. The interaction between time and program was included because previous research suggests that the effects of different programs on the rate of English acquisition may change over time (Thompson, 2017; Umansky & Reardon, 2014). As shown in Table 36, the two models yielded similar results. Moreover, the BIC of Model ProgramTI was much smaller than that of Model ProgramTV, which suggests that Model ProgramTI provided a better fit to the data.

Table 36. Results for two program models, Model ProgramTV and Model ProgramTI, coding instructional program as time-varying and time-invariant, respectively.

Variable	Model ProgramTV (program coded as time-varying)				Model ProgramTI (program coded as time-invariant)			
	Estimate	SE	z value	p	Estimate	SE	z value	p
time0	-3.16	0.02	-138.37	<.001	-3.15	0.02	-138.64	<.001
time1	-4.01	0.04	-103.74	<.001	-3.98	0.04	-103.01	<.001
time2	-2.77	0.02	-110.92	<.001	-2.74	0.03	-109.32	<.001
time3	-1.78	0.02	-87.61	<.001	-1.75	0.02	-85.49	<.001
time4	-0.69	0.02	-33.86	<.001	-0.68	0.02	-32.46	<.001
time5	-0.73	0.03	-21.17	<.001	-0.77	0.04	-21.11	<.001
TB	-1.05	0.11	-9.31	<.001	-1.26	0.10	-12.54	<.001
DB	-2.07	0.23	-9.11	<.001	-2.10	0.22	-9.40	<.001
NP	0.40	0.16	2.54	0.011	0.38	0.18	2.12	0.0345
TB*No. of years	0.17	0.03	4.94	<.001	0.28	0.02	10.20	<.001
DB*No. of years	0.36	0.06	5.82	<.001	0.38	0.06	6.35	<.001
NP*No. of years	0.03	0.05	0.59	0.553	0.06	0.07	0.98	0.329
Deviance (df)	73607 (161199)				73562 (161199)			
AIC	73631				73586			
BIC	73751				73706			

Note. Time0 represents the year when ELs entered school in kindergarten. Time1 through time5 represent the first through fifth years of school instruction since ELs entered kindergarten, respectively.

## B.4 Poverty Status

Unlike previous EL research where poverty status was treated as time-invariant, this study accounted for possible changes in ELs' poverty status over time by coding poverty as a time-varying variable. I compared three popular methods for coding time-varying variables, all displayed in Table 37. The first method is denoted *time-specific* coding, whereby poverty status was coded dichotomously based on whether a child received free/reduced-price lunch in a given time period. With this method, each *Yes* was coded as 1 and *No* coded as 0 (see Table 37). The second method led to a dichotomous poverty indicator called *ever in poverty*, which represents whether a child ever received a subsidized meal between school entry and a given, later time period. The variable was coded 0 until the first time when the student was ever in poverty and was coded 1 thereafter (see Table 37). Unlike the first two methods, the third method produced a continuous poverty indicator, which records the total number of years in which the student received subsidized meals between school entry and a given, later time period. For each student, their value on this variable accumulated by 1 with each additional year in which they were in poverty (see Table 37).

Table 37. Three methods for coding poverty as a time-varying variable (2 fake examples)

Fake case	Time	Free/reduced-price lunch	Coding scheme		
			1. Time-specific poverty status	2. Ever in poverty	3. Number of years in poverty
ID 1	0	No	0	0	0
	1	Yes	1	1	1
	2	Yes	1	1	2
	3	Yes	1	1	3
	4	No	0	1	3
	5	No	0	1	3
ID 2	0	Yes	1	1	1
	1	Yes	1	1	2
	2	Yes	1	1	3
	3	No	0	1	4
	4	No	0	1	4
	5	No	0	1	4

Note. Time0 represents the year when ELs entered school in kindergarten. Time1 through time5 represent the first through fifth years of school instruction since ELs entered kindergarten, respectively.

To determine which of the three methods was optimal for coding poverty, I tested the statistical efficiency of the three time-varying poverty variables in predicting the outcome of English proficiency attainment along with time (categorical variable). The results for the three poverty models are shown in Table 38. Because the model including *ever in poverty* had the lowest BIC among the three models, the second method was selected to code poverty status.

Table 38. Results for three models that used poverty status and time to predict the outcome of proficiency attainment, each using a different method to code poverty status.

Variable	Model Poverty1 (Time-specific)			Model Poverty2 (Ever in poverty)			Model Poverty3 (Number of years in poverty)		
	EST.	SE	z value	EST.	SE	z value	EST.	SE	z value
time0	-2.54	0.02	-103.55***	-2.48	0.02	-101.19***	-3.03	0.02	-132.38***
time1	-3.33	0.04	-83.39***	-3.22	0.04	-80.12***	-3.65	0.04	-92.73***
time2	-2.05	0.03	-74.42***	-1.90	0.03	-67.34***	-2.20	0.03	-79.82***
time3	-0.98	0.02	-40.17***	-0.80	0.03	-31.21***	-0.97	0.03	-36.92***
time4	0.24	0.03	8.96***	0.44	0.03	15.4***	0.47	0.03	14.28***
time5	0.24	0.04	6.23***	0.45	0.04	11.27***	0.74	0.05	15.48***
Poverty	-1.19	0.02	-54.78***	-1.33	0.02	-58.31***	-0.30	0.01	-45.81***
Deviance (df)	71107 (161204)			70736 (161204)			71952 (161204)		
AIC	71121			70750			71966		
BIC	71191			70820			72036		

Note. Time0 represents the year when ELs entered school in kindergarten. Time1 through time5 represent the first through fifth years of school instruction since ELs entered kindergarten, respectively; EST. = ESTIMATE; \*\*\*  $p < .001$ , \*\*  $p < .005$ , \*  $p < .01$ .

## B.5 English use at home

In this study, English use at home was treated as a time-varying variable. I selected the best method for coding English use at home following the same procedure as I did for poverty status. Specifically, I created three home-English-use variables using each of the three methods mentioned above and compared their efficiency in predicting the outcome of English proficiency attainment through statistical modeling. Table 39 illustrates how I implemented the three coding methods. *Time-specific use of English at home* represents whether the student used English at home in a given time period. *Ever using English at home* represents whether the student ever used English at home between school entry and a given, later time period. *Number of years using English at home* is a measure of the total number of years that the student used English at home between school entry and a given, later time period.

Table 39. Three methods for coding English use at home as a time-varying variable (2 fake examples)

Fake case	Time	English use at home	Coding scheme		
			1. Time-specific use of English at home	2. Ever using English at home	3. Number of years using English at home
ID 1	0	No	0	0	0
	1	Yes	1	1	1
	2	Yes	1	1	2
	3	Yes	1	1	3
	4	No	0	1	3
	5	No	0	1	3
ID 2	0	Yes	1	1	1
	1	Yes	1	1	2
	2	Yes	1	1	3
	3	No	0	1	4
	4	No	0	1	4
	5	No	0	1	4

*Note.* Time0 represents the year when ELs entered school in kindergarten. Time1 through time5 represent the first through fifth years of school instruction since ELs entered kindergarten, respectively.

I tested three models that used English use at home and time (categorical variable) to predict the outcome of proficiency attainment, with English use at home coded as time-specific use of English at home (Model HomeEnglish1), ever using English at home (Model HomeEnglish1), and number of years using English at home (Model HomeEnglish1), respectively. Table 40 displays the model results. Among the three models, Model HomeEnglish1 demonstrated the best fit to the data (as evidenced the BIC), followed closely by Model HomeEnglish2, whereas Model HomeEnglish3 fit the data the worst. Although the BIC of Model HomeEnglish1 was slightly lower than that of Model HomeEnglish2 (delta BIC = 16), the two models yielded practically the same results with regard to parameter estimates, standard errors, and z values. Given these findings, I decided to use the second coding method because it represents the original English-use-at-home data with a much smaller number of data patterns than the first method, which thus makes it easier to interpret and visualize the estimated effect. To illustrate my point, one can think about a randomly selected student who had data for all six time periods (time0 through time5). No matter how the student's original home-English-use data changed over the years, the coded data would fall into one of the following seven ever-using-English-at-home-pattens (each digit representing the coded value in a given time period from time0 to time5, ordered left to right):

- 000000 (never used English at home),
- 000001 (used English only in the last time period),
- 000011 (used English only in the last two time periods),
- 000111 (used English only in the last three time periods),
- 001111 (used English only in the last four time periods),
- 011111 (used English only in the last five time periods),

- 111111 (used English in all six time periods).

If the data were coded using the first method, the number of possible patterns would be much higher ( $2^6 = 64$ ).

Table 40. Results for three models that used English use at home and time to predict the outcome of proficiency attainment, each using a different method to code English use at home.

Variable	Model HomeEnglish1 (Time-specific use of English at home)			Model HomeEnglish2 (Ever using English at home)			Model HomeEnglish3 (Number of years using English at home)		
	EST.	SE	z value	EST.	SE	z value	EST.	SE	z value
time0	-3.27	0.02	-143.15***	-3.27	0.02	-143.23***	-3.24	0.02	-144.00***
time1	-4.11	0.04	-105.92***	-4.11	0.04	-105.85***	-4.08	0.04	-105.73***
time2	-2.86	0.03	-113.14***	-2.86	0.03	-112.87***	-2.84	0.03	-113.13***
time3	-1.86	0.02	-90.31***	-1.86	0.02	-89.89***	-1.85	0.02	-89.50***
time4	-0.76	0.02	-37.48***	-0.77	0.02	-37.36***	-0.76	0.02	-36.62***
time5	-0.79	0.03	-23.21***	-0.80	0.03	-23.24***	-0.79	0.03	-22.94***
English use at home	0.29	0.03	11.39***	0.26	0.02	10.58***	0.07	0.01	7.618***
Devian- ce (df)	73915 (161204)			73931 (161204)			73983 (161204)		
AIC	73929			73945			73997		
BIC	73999			74015			74067		

Note. Time0 represents the year when ELs entered school in kindergarten. Time1 through time5 represent the first through fifth years of school instruction since ELs entered kindergarten, respectively; EST. = ESTIMATE; \*\*\*  $p < .001$ , \*\*  $p < .005$ , \*  $p < .01$ .

## B.6 Retention

Retention was treated as a time-varying variable in this study. I compared two methods for coding retention. The first method resulted in a variable called *time-specific retention status*, which indicates whether the student was retained in a given time period. The second method resulted in a variable called *ever retained*, which represents whether the student was ever retained between school entry and a particular, later time period. Table 41 illustrates how I implemented the two coding methods. Statistical analyses showed that *ever retained* had stronger statistical efficiency than *time-specific retention status* when used to predict the outcome of proficiency attainment along with time (categorical variable; see Table 42 for model results). I therefore decided to use the second method to code retention.

Table 41. Two methods for coding retention as a time-varying variable (2 fake examples)

Fake case	Time	Retention	Coding scheme	
			1. Time-specific retention status	2. Ever retained
ID 1	0	No	0	0
	1	Yes	1	1
	2	Yes	1	1
	3	Yes	1	1
	4	No	0	1
	5	No	0	1
ID 2	0	Yes	1	1
	1	Yes	1	1
	2	Yes	1	1
	3	No	0	1
	4	No	0	1
	5	No	0	1

*Note.* Time0 represents the year when ELs entered school in kindergarten. Time1 through time5 represent the first through fifth years of school instruction since ELs entered kindergarten, respectively.

Table 42. Results for two models that used retention and time to predict the outcome of proficiency attainment, each using a different method to code retention.

Variable	Model Retention1 (Time-specific retention)			Model Retention2 (Ever retained)		
	EST.	SE	z value	EST.	SE	z value
time0	-3.23	0.02	-143.82***	-3.23	0.02	-143.82***
time1	-4.12	0.04	-103.47***	-4.01	0.04	-103.97***
time2	-2.83	0.02	-113.40***	-2.75	0.02	-110.46***
time3	-1.82	0.02	-90.64***	-1.74	0.02	-86.06***
time4	-0.72	0.02	-36.10***	-0.63	0.02	-31.07***
time5	-0.75	0.03	-22.00***	-0.63	0.03	-18.37***
Retention	0.56	0.07	7.54***	-0.93	0.05	-19.82***
Deviance	73989			73557		
(df)	(161204)			(161204)		
AIC	74003			73571		
BIC	74073			73641		

*Note.* Time0 represents the year when ELs entered school in kindergarten. Time1 through time5 represent the first through fifth years of school instruction since ELs entered kindergarten, respectively; EST. = ESTIMATE; \*\*\*  $p < .001$ , \*\*  $p < .005$ , \*  $p < .01$ .

## APPENDIX C Demographic Characteristics of the Subsample

This section presents the mean demographic characteristics for the subsample by cohort. As described in Chapter 7, the subsample consists of those ELs in the three older cohorts (2013-2014 kindergarten cohort, 2014-2015 kindergarten cohort, and 2015-2016 kindergarten cohort) who had not attained English proficiency by the end of the first year of instruction. As shown in Table 43, the demographic characteristics of the subsample are very similar to those of the full sample (see Chapter 7 Table 2).

Table 43. Mean demographic characteristics of the subsample by cohort

Student characteristic	Total subsample (N = 27,047)		K13-14 <sup>a</sup> (n = 9,345)		K14-15 (n = 8,768)		K15-16 (n = 8,934)	
Gender	n	%	n	%	n	%	n	%
Female	13063	48%	4453	48%	4316	49%	4294	48%
Male	13984	52%	4892	52%	4452	51%	4640	52%
Race								
Hispanic or Latino	10143	38%	3624	39%	3184	36%	3335	37%
African-American or Black	720	3%	250	3%	227	3%	243	3%
AIAN/NHPI <sup>b</sup>	77	0%	24	0%	29	0%	24	0%
Asian	5803	21%	1867	20%	1920	22%	2016	23%
White	10006	37%	3485	37%	3308	38%	3212	36%
Two or more races	298	1%	95	1%	100	1%	103	1%
Primary disability type								
None	23458	87%	7996	86%	7656	87%	7806	87%
Learning disabilities	814	3%	382	4%	243	3%	189	2%
Speech and language disabilities	1970	7%	685	7%	603	7%	682	8%
Other disabilities	805	3%	282	3%	266	3%	257	3%
Primary home language (family) <sup>c</sup>								
Spanish	10324	38%	3696	40%	3250	37%	3378	38%
Arabic	6954	26%	2303	25%	2319	26%	2332	26%
Bengali	673	2%	186	2%	208	2%	279	3%
Chinese	752	3%	265	3%	244	3%	243	3%
Albanian	571	2%	212	2%	201	2%	158	2%
AfroAsiatic	816	3%	316	3%	263	3%	237	3%
Austronesian	743	3%	283	3%	229	3%	231	3%
BaltoSlavic	764	3%	269	3%	257	3%	238	3%
Dravidian	955	4%	274	3%	318	4%	363	4%
IndoIranian	1445	5%	478	5%	492	6%	475	5%
Italic/Germanic	790	3%	298	3%	243	3%	249	3%



Table 43 cont'd

Japanese/Korean	784	3%	264	3%	273	3%	247	3%
Other	1476	5%	501	5%	471	5%	504	6%
Primary home language								
Spanish	10322	38%	3695	40%	3250	37%	3377	38%
Arabic	6959	26%	2302	25%	2324	27%	2333	26%
Bengali	675	2%	186	2%	210	2%	279	3%
Chinese	753	3%	265	3%	244	3%	244	3%
Albanian	570	2%	211	2%	201	2%	158	2%
Aramaic	365	1%	122	1%	116	1%	127	1%
Japanese	479	2%	157	2%	178	2%	144	2%
Telugu	502	2%	138	1%	174	2%	190	2%
Urdu	428	2%	145	2%	156	2%	127	1%
Vietnamese	536	2%	207	2%	165	2%	164	2%
Other	5458	20%	1917	21%	1750	20%	1791	20%
Poverty								
Never poor	6236	23%	2156	23%	2004	23%	2076	23%
Ever poor	20811	77%	7189	77%	6764	77%	6858	77%
Program <sup>d</sup>								
Ever in transitional bilingual programs	3034	11%	1865	20%	711	8%	458	5%
Ever in dual-language bilingual programs	1161	4%	446	5%	350	4%	365	4%
Ever had no program	783	3%	351	4%	244	3%	188	2%
Only ESL program	22179	82%	6754	72%	7490	85%	7935	89%
Home English use								
Ever using English at home	5044	19%	1663	18%	1623	19%	1758	20%
Never using English at home	22003	81%	7682	82%	7145	81%	7176	80%
Retention <sup>e</sup>								
Ever retained	2502	9%	791	8%	791	9%	920	10%
Never retained	24545	91%	8554	92%	7977	91%	8014	90%
Ten largest districts								
Dearborn City School	2946	11%	1020	11%	961	11%	965	11%
Detroit Public Schools								
Community	1476	5%	469	5%	468	5%	539	6%
Grand Rapids Public Schools	1262	5%	467	5%	389	4%	406	5%
Warren Consolidated Schools	1219	5%	394	4%	341	4%	484	5%
Utica Community Schools	1111	4%	409	4%	383	4%	319	4%
Troy School	913	3%	300	3%	289	3%	324	4%
Ann Arbor Public Schools	786	3%	259	3%	282	3%	245	3%
Farmington Public School	578	2%	211	2%	191	2%	176	2%

Table 43 cont'd

Kentwood Public Schools	551	2%	176	2%	137	2%	238	3%
Plymouth-Canton								
Community Schools	548	2%	180	2%	190	2%	178	2%

*Note.* <sup>a</sup>K13-14 = 2013-2014 kindergartener cohort, K14-15 = 2014-2015 kindergartener cohort, K15-16 = 2015-2016 kindergartener cohort.

<sup>b</sup>AIAN/NHPI = American Indian or Alaska Native/Native Hawaiian or Other Pacific Islander.

<sup>c</sup>The counts of Spanish, Arabic, Bengali, Chinese, and Albanian speakers differ slightly between primary home language (family) and primary home language coding because some students reported different home languages across the years, and for those students who reported to speak two or more languages equally likely at home, a random language was chosen as their home language.

<sup>d</sup>The percentages of different program categories do not add up to 100% because there was some overlap among students who were ever in transitional bilingual programs, who were ever in dual-language bilingual programs, and who ever had no program.

<sup>e</sup>2487 students were retained once in grade, 15 were retained twice in grade.

## APPENDIX D Results for Multicollinearity Checks

I checked if there was multicollinearity among the six predictors of interest: primary disability type, primary home language, poverty status, home English use, instructional programming, and retention. Crosstabulation was used to check the pairwise relationships among three nominal predictors: primary disability type, primary home language, and instructional programming, as well as their relationships with each of the three discretized, ordinal predictors: poverty status, home English use, and retention. Pairwise polychoric correlation was calculated for the three discretized, ordinal predictors (see Table 44).

Table 44. Polychoric correlation among poverty status, home English use, instructional programming, and retention

	Home English use	Poverty status	Retention
Home English use	1		
Poverty status	-0.152	1	
Retention	0.003	0.183	1

These analyses indicated that only two predictors, primary home language and poverty status, had noticeable correspondence, and their relationship is illustrated in Table 45.

Table 45. Relationship between primary home language and poverty status

Primary home language	time1 (Year2)	Proportion of students ever in poverty			
		time2 (Year3)	time3 (Year4)	time4 (Year5)	time5 (Year6)
Spanish	90.3%	92.9%	94.4%	96.0%	97.8%
AfroAsiatic	84.2%	86.3%	87.6%	89.7%	89.6%
Albanian	65.6%	71.2%	75.6%	80.3%	82.1%
Arabic	90.2%	92.1%	93.9%	95.5%	97.0%
Austronesian	52.5%	58.0%	62.3%	68.6%	68.6%
BaltoSlavic	36.5%	42.1%	49.6%	56.2%	45.3%
Bengali	88.6%	90.9%	92.3%	92.7%	93.4%
Chinese	30.3%	34.7%	40.1%	45.9%	52.8%
Dravidian	2.0%	2.6%	3.3%	5.3%	7.3%
IndoIranian	33.3%	37.0%	43.7%	54.7%	59.1%
ItalicGermanic	36.7%	41.4%	47.8%	58.0%	64.6%
JPN-KOR	10.2%	11.8%	17.5%	23.9%	35.0%
Other	65.2%	68.7%	73.2%	80.2%	85.6%

Specifically, students from some home language backgrounds (e.g., Spanish, Arabic, Bengali, AfroAsiatic) were more likely to have ever lived in poverty during the study period than those

from some other backgrounds (e.g., Chinese, Dravidian, Japanese, Korean). For example, over 80% of children speaking Spanish, Arabic, Bengali, or AfroAsiatic languages as their home language had ever lived in poverty in each time period from time 1 through 6, compared to 10% of children speaking Dravidian languages and 30% of children speaking Chinese, Japanese, or Korean. Although the correspondence between primary home language and poverty status was moderately strong, such correspondence should not cause any multicollinearity problem in the modeling process.

Here are two additional notes about the multicollinearity checks described above: First, home English use, poverty status, and retention were operationalized as three time-varying variables indicating whether the student ever used English at home, ever lived in poverty, or was ever retained in grade during the study period; Second, pairwise polychoric correlation was calculated using data in the long format (a.k.a. person-period dataset, i.e., each student had one row of data per time period).

## APPENDIX E Assumption Testing

In this section, I describe how I tested two important assumptions for survival analysis: the assumption of non-informative censoring and the proportionality assumption.

### E.1 Non-informative Censoring

In this study, students may be censored through two mechanisms: (a) students had not attained proficiency by the end of data collection (i.e., 2018-2019); (b) students left the study early and had not attained proficiency before they left. Rather than removing censored students, survival analysis allows those students to remain in the risk set (the group of students who have not attained proficiency but are eligible to do so) and to contribute to the estimated risk of English proficiency (or other outcome) attainment in each time period they were observed. Table 46 displays the number of students in the *full sample* who were censored due to mechanisms (a) and (b) by time period (time0 through time5).

Table 46. The number of students in the full sample who were censored due to mechanisms (a) and (b) by time period

Time	Starting number	Event = 1 (attained proficiency)	Censored due to mechanism (a)	Censored due to mechanism (b)	Total censored	Total censored proportion
0	54146	2062	7950	3630	11580	21.39%
1	40504	686	7548	1838	9386	23.17%
2	30432	1724	7034	1026	8060	26.49%
3	20648	2898	5792	466	6258	30.31%
4	11492	3776	3592	135	3727	32.43%
5	3989	1284	2705	0	2705	67.81%
<b>Total</b>		<b>12430</b>	<b>34621</b>	<b>7095</b>	<b>41716</b>	

*Note.* Time0 represents the year when ELs entered school in kindergarten. Time1 through time5 represent the first through fifth years of school instruction since ELs entered kindergarten, respectively.

For survival analysis to be valid, censoring must be non-informative (Thompson, 2017). That is, censoring must be “independent of event occurrence and the risk of event occurrence” (Singler & Willet, 2003, p. 318). In this study, I assume that all censoring is non-informative and is unrelated to students’ likelihood of attaining English proficiency. In other words, I assume ELs in a given risk set are representative of the population of ELs who would have entered the risk set if they had not been censored. By design, this assumption most reasonably holds for censoring due to the end of data collection (mechanism [a]) because the source of censoring in this situation is the researcher’s data collection schedule. Of the 41,716 students in the full sample who were censored, a majority (82.99%) was censored because data collection ended, not because of actions taken by the students or their parents.

While I also assume non-informative censoring for censored cases ( $n = 7095$ ) that left the study prior to the end of data collection (mechanism [b]), this assumption may bias the findings

to some degree if those censored ELs were more or less likely to attain proficiency in the periods after they left, compared to students who remained in the dataset during those periods. Thompson (2017) argued that lack of relationship between study covariates and attrition serves as useful evidence for non-informative censoring. Following Thompson, I conducted a discrete-time hazard model to predict attrition using ACCESS scores and the following covariates: gender, race, age, primary disability type, primary home language, initial program in kindergarten, poverty status, English use at home, retention, and cohort. Results of the analysis suggested that significant relationships existed between attrition and many of the covariates (see Table 47 for a summary of the results for significant covariates).

Table 47. Summary of the results of the discrete-time hazard model predicting attrition

Comparison	A	B	C
1. Gender	Males	Females	
2. Race	Non-Hispanic	Hispanic	
3. Age	Students who were older upon school entry	Students who were younger upon school entry	
4. Disability type	Students who were identified with disabilities other than specific learning disabilities or speech or language impairments	Students with no disabilities	Students with specific learning disabilities or speech or language impairments
5. Primary home language/family	Students who spoke Chinese, Japanese, Korean, or a home language of the Italic or Germanic families	Spanish-speaking students	Students who spoke Arabic, Albanian, Bengali, or a home language of the AfroAsiatic, Austroneisan, or BaltoSlavic families
6. Poverty	Students who were never in poverty	Students who were ever in poverty	
7. Retention	Students who were ever retained	Students who were never retained	
8. Test scores	Students who attained lower scores on ACCESS	Students who attained higher scores on ACCESS	

*Note.* From left to right, A was significantly *more* likely to leave the dataset early than B, and B was significantly *more* likely to leave the dataset early than C.

Because those covariates that significantly predicted attrition were found to differentially associate with students' likelihood of attaining proficiency (see Appendix H for reports on covariate effects), the direction of the bias resulting from attrition is not entirely clear. For example, estimates of time to proficiency may contain negative bias because students receiving free/reduced-price lunch were significantly less likely to leave the dataset early and were significantly less likely to attain proficiency; however, positive bias may also exist given that

students who attained higher scores on ACCESS were less likely to leave the dataset early. Although the analysis suggested that estimates of time to proficiency may contain both positive and negative bias caused by attrition-related censoring, the bias is likely to be trivial because only a small proportion of students were censored due to attrition (mechanism [b]).

## E.2 Proportionality Assumption

The second key assumption for survival analysis is the proportionality assumption. This assumption stipulates that the effect of a predictor on the outcome is identical in every time period. I tested if this assumption was true for each of the following predictors: primary disability type, primary home language, instructional programming, poverty status, English use at home, and retention. The testing procedure I used was recommended by Singler and Willet (2003). It involved comparing the goodness-of-fit of two discrete-time hazard models for each predictor, with one model assuming that the predictor's effect on the outcome was invariant in every time period (Model 1), and the other model allowing the effect of the predictor to vary across time (Model 2). The general function forms of two hazard models are given as follows:

$$\text{Logit } h(t_{ij}) = [\alpha_1 D_{i1} + \alpha_2 D_{i2} + \dots + \alpha_j D_{ij}] + \beta_1 P_{ij} \quad (\text{Model 1})$$

$$\text{Logit } h(t_{ij}) = [\alpha_1 D_{i1} + \alpha_2 D_{i2} + \dots + \alpha_j D_{ij}] + \beta_1 P_{ij} + \beta_2 P_{ij} * T_{ij} \quad (\text{Model 2})$$

where  $D_1$  through  $D_j$  represent a series of dummy-coded indicators for each time period in which the student was observed,  $P_{ij}$  is a predictor of interest,  $T_{ij}$  is a continuous time variable representing the number of years that has elapsed since the student entered kindergarten, and  $P_{ij} * T_{ij}$  represents an interaction between the predictor and time. Because Model 1 does not contain the interaction term, it assumes that the effect of the predictor does not depend on time (operationalized as the number of years the student has been enrolled in Michigan public schools). By including the interaction term, Model 2 allows the effect of the predictor to change in a linear fashion over time (Singler & Willet, 2003). I focused only on the primary outcome of English proficiency attainment in assumption testing.

Singler and Willet (20103) provided the following principles to evaluate the proportionality assumption: If Model 2 is sufficiently superior to Model 1, one should relax the proportionality assumption and adopt Model 2 because the assumption is violated; if Model 2 does not demonstrate sufficient improvement in fit compared to Model 1, one should adopt Model 1 for parsimony concerns. While Singler and Willet recommended researchers to use a likelihood-ratio test (LRT) to compare the two models, I decided to use BIC instead because significance testing, which underpins LRTs, is sensitive to sample size and tends to be significant regardless of the size of the effect when the sample size is large (Armstrong, 2019).

Table 48 displays the BIC values of Models 1 and 2 for each of the six predictors of interest. As one can see, the proportionality assumption only holds for primary disability type (as evidenced by a smaller BIC for Model 1). For the other five predictors, the interaction term between the predictor and time is needed because the proportionality assumption does not appear to be plausible.

Table 48. BIC of Models 1 and Model 2 for six predictors of interest

Predictor	BIC of Model 1	BIC of Model 2	Superior model
Primary disability type	72478	72483	Mode 1
Primary home language	70184	70080	Model 2
Program	73833	73706	Model 2
Poverty status	70820	70735	Model 2
English use at home	74015	73994	Model 2
Retention	73641	73334	Model 2



## APPENDIX F. Results for Research Question 1 Based on the Subsample

In this section, I present results for the first research question based on data of the subsample. The first research question asks about the typically length of time it takes for English learner (EL) children in Michigan public schools to attain proficiency. The subsample consists of those students in the three older cohorts (2013-2014 kindergarten cohort, 2014-2015 kindergarten cohort, and 2015-2016 kindergarten cohort) who had not attained proficiency by the end of the first year of school (kindergarten).

Table 49 is a life table showing the subsample's history of proficiency attainment from the beginning of time (second year in school, time 1) through the end of data collection (sixth year in school, time 5). Specifically, this table summarizes the number of students present in the risk set at the beginning of each time period (column C), the number of students who attained proficiency during that time period (column D), and the number of students who were censored by the end of that time period (column E). Using these basic counts, the table also displays the sample's hazard of reaching proficiency in each time period (column F), which is equal to the proportion students who attained proficiency during that period.

Table 49. Life table for proficiency attainment of ELs in the sub sample, using data from 2014-2015 through 2018-2019

A	B	C	D	E	F
Time	Years since kindergarten entry	Beginning total (in the risk set)	Attained proficiency	Total censored	Hazard
1	2	24702	377	1356	1.53%
2	3	22969	1295	1026	5.64%
3	4	20648	2898	6258	14.04%
4	5	11492	3776	3727	32.86%
5	6	3989	1284	2705	32.19%
<b>Total</b>			<b>9630</b>	<b>15072</b>	

*Note.* A risk set (column C) consists of students who had not attained proficiency by the beginning of a particular time period but were eligible to do so during that time period. Students are censored if they left the study prior to the end of a particular time period without attaining proficiency (column E). Hazard (column F) is equal to the proportion of students who began a given time period in the risk set and subsequently left the risk set because they attained proficiency during that time period.

Table 50 displays the point hazard estimates and their 95% confidence intervals (CIs) derived from the unconditional discrete-time hazard model (Equation 1 in Chapter 7). The table also includes the survival probability, or the cumulative proportion of students who have not reached proficiency, and the cumulative proportion of proficient students. By design, the hazard estimates based on the unconditional model are the same as those estimates presented in the life table (because time is the only model covariate).

Table 50. Results of the unconditional discrete-time hazard model (specified in Equation 1) predicting hazard of proficiency attainment as a function of time

Time	Years since school entry	Logit hazard	SE (for logit hazard)	Hazard	95% CI for hazard		Survival	Cumulative proportion of proficient students
					Upper	Lower		
1	2	-4.1670	0.0519	0.0153	0.0138	0.0169	0.9847	0.0153
2	3	-2.8176	0.0286	0.0564	0.0535	0.0594	0.9292	0.0708
3	4	-1.8124	0.0200	0.1404	0.1357	0.1452	0.7988	0.2012
4	5	-0.7146	0.0199	0.3286	0.3200	0.3372	0.5363	0.4637
5	6	-0.7451	0.0339	0.3219	0.3076	0.3366	0.3637	0.6363

*Note.* Hazard is equal to the proportion of students who began a given time period in the risk set and subsequently left the risk set because they attained proficiency during that time period. Survival probability is equal to the cumulative proportion of students who had not attained proficiency by the end of a given time period. The proportions in the last two columns account for students who were censored.

Examining changes in the estimated hazards (column 3 in Table 50) across time suggests that the subsample's likelihood of reaching proficiency increased rapidly during elementary grades, peaking after the students spent five or six years in Michigan public schools (fourth or fifth grade if the students had not been retained). The median time to proficiency, as one can extrapolate from the cumulative proficiency rates (last column of Table 50), was around five years. These results are similar to the results obtained for the full sample.

## APPENDIX G. Model Results for Research Question 2 Based on the Full Sample

In this section, I present the results for research question 2 based on the full sample. Research question 2 asks about the effects of the focal predictors of interest (i.e., primary disability status, primary home language, instructional programming, poverty status, home English use, and retention) on students' likelihood of attaining proficiency. I answered this question by fitting the conditional discrete-time hazard model as specified in Equation 3 in Chapter 7. As I note in Chapter 7, the conditional hazard model fitted to the *subsample* was different from the model fitted to the *full sample* in that the *subsample* model included ELs' initial proficiency (end-of-kindergarten ACCESS scores) and its interaction with time as two control variables, but these control variables could not be included in the *full-sample* model due to a multicollinearity problem.

Table 51 displays the model results for the full sample. For clarity considerations, the table only includes the six predictors and their interactions with time. In the subsample model, initial proficiency had a positive, significant main effect on students' likelihood of attaining proficiency as well as a negative interaction with time. In comparing the model results of the full sample to those of the subsample (Table 26 in Chapter 8), one can see that the coefficients (in logits) of several predictors are different between the two models, suggesting that initial proficiency may be confounded with those predictors. (I consider a predictor's coefficients "different" between the two models if the absolute value of the difference is at least twice as large as the predictor's standard error.) All noticeable differences between the two models are listed as follows:

- 1) The negative effects of the three *primary disability* predictors (i.e., specific learning disabilities, speech/language impairments, and other disabilities) were all smaller (in absolute value) in the subsample model than in the full-sample model. This is perhaps because students with disabilities tended to receive lower ACCESS scores in kindergarten than their peers without disabilities; hence, the effects of the disability predictors were partially explained by initial proficiency in the subsample model.
- 2) Similarly, the main effects of *poverty status* and *dual-language program* were smaller (less negative) in the subsample model than in the full sample model maybe because students with low initial English proficiency were more likely to live in poverty and/or to be placed in dual-language bilingual programs.
- 3) In contrast, the main effect of *retention* was larger (more positive) in the subsample model than in the full sample model. One possible explanation is that compared with students who were never retained, their peers who were ever retained were more likely to have low initial proficiency and were more likely to attain proficiency in early elementary grades. Therefore, when examining the retention effect within the same level of initial proficiency (as the subsample model did), the effect of retention would appear to be larger.
- 4) The only *primary-home-language* predictor that demonstrated substantially different effects between the two models was Japanese/Korean. In the subsample model, Japanese/Korean had a large main effect (more positive) and a significant, negative interaction effect, whereas in the full-sample model, it had a small main effect (less positive) and an insignificant interaction with time. This is probably because Japanese/Korean ELs were more likely to receive lower initial proficiency scores than

Spanish ELs but were more likely to attain proficiency than Spanish learners in early elementary grades.

Table 51. Results for the conditional discrete-time hazard model predicting proficiency hazard for students in the full sample, using data from 2013-2014 through 2018-2019

	Logit	SE (for logit)	z value	p value	OR	95% CIs for OR	
						Lower	Upper
Primary disability type (reference = no disabilities)							
Specific learning disabilities	-2.987	0.166	-18.03	<.001***	0.05	0.04	0.07
Speech/language impairments	-0.929	0.052	-17.94	<.001***	0.40	0.36	0.44
Other disabilities	-1.793	0.114	-15.72	<.001***	0.17	0.13	0.21
Primary home language (reference = Spanish)							
AfroAsiatic	-0.097	0.182	-0.54	.592	0.91	0.64	1.30
Albanian	0.289	0.179	1.62	.106	1.33	0.94	1.90
Arabic	0.209	0.119	1.75	.079	1.23	0.98	1.56
Austronesian	0.456	0.159	2.87	.004**	1.58	1.16	2.15
BaltoSlavic	0.361	0.156	2.32	.020*	1.43	1.06	1.95
Bengali	0.708	0.170	4.17	<.001***	2.03	1.46	2.83
Chinese	0.937	0.145	6.45	<.001***	2.55	1.92	3.39
Dravidian	1.242	0.136	9.12	<.001***	3.46	2.65	4.52
IndoIranian	0.948	0.134	7.08	<.001***	2.58	1.99	3.35
Italic/Germanic	0.464	0.140	3.31	<.001***	1.59	1.21	2.09
Japanese/Korean	0.421	0.151	2.78	.005**	1.52	1.13	2.05
Other home languages	0.471	0.136	3.47	<.001***	1.60	1.23	2.09
Instructional program (reference = ESL)							
Transitional bilingual	-0.268	0.113	-2.38	.017*	0.77	0.61	0.95
Dual-language bilingual	-1.018	0.247	-4.12	<.001***	0.36	0.22	0.59
No program	0.416	0.187	2.23	.026*	1.52	1.05	2.19
Poverty status (reference = never in poverty)							
Retention	-0.759	0.056	-13.64	<.001***	0.47	0.42	0.52
(reference = never retained)							
Home English use	1.898	0.121	15.64	<.001***	6.67	5.26	8.47
(reference = never used English at home)							
English at home)	0.318	0.052	6.12	<.001***	1.37	1.24	1.52

Table 51 cont'd

Interactions							
AfroAsiatic*time	-0.076	0.046	-1.66	.097	0.93	0.85	1.01
Albanian*time	-0.050	0.049	-1.03	.305	0.95	0.86	1.05
Arabic*time	-0.133	0.021	-6.30	<.001***	0.88	0.84	0.91
Austronesian*time	0.011	0.042	0.26	.795	1.01	0.93	1.10
BaltoSlavic*time	-0.066	0.042	-1.56	.119	0.94	0.86	1.02
Bengali*time	-0.100	0.043	-2.32	.021*	0.90	0.83	0.98
Chinese*time	-0.085	0.041	-2.07	.039*	0.92	0.85	1.00
Dravidian*time	-0.349	0.035	-9.83	<.001***	0.71	0.66	0.76
IndoIranian*time	-0.274	0.031	-8.95	<.001***	0.76	0.72	0.81
Italic/Germanic*time	-0.090	0.042	-2.17	.030*	0.91	0.84	0.99
Japanese/Korean*time	0.011	0.048	0.23	.819	1.01	0.92	1.11
Other home languages*time	-0.136	0.033	-4.16	<.001***	0.87	0.82	0.93
Transitional bilingual*time	0.044	0.030	1.48	.139	1.04	0.99	1.11
Dual-language bilingual*time	0.230	0.067	3.45	<.001***	1.26	1.10	1.43
No program*time	-0.009	0.071	-0.13	.899	0.99	0.86	1.14
Poverty status*time	0.054	0.020	2.68	.007**	1.06	1.01	1.10
Retention*time	-0.809	0.040	-20.19	<.001***	0.45	0.41	0.48
Home English use*time	-0.075	0.017	-4.30	<.001***	0.93	0.90	0.96

## APPENDIX H. Full Model Results for Research Question 2 Based on the Full Sample

In this section, I present the full results for fitting the conditional discrete-time hazard model (Equation 3 in Chapter 7) to the data of the subsample (see Table 52).

Table 52. Full results for the conditional discrete-time hazard model predicting proficiency hazard for students in the subsample, using data from 2013-2014 through 2018-2019

	Estimate	Std. Error	z value	p value	
time1	-5.34	0.144	-37.21	<.001	***
time2	-3.45	0.108	-32.02	<.001	***
time3	-1.86	0.091	-20.39	<.001	***
time4	-0.02	0.097	-0.23	0.818	
time5	0.52	0.123	4.23	<.001	***
female	0.41	0.026	15.81	<.001	***
Race (reference = Hispanic/Latino)					
African-American or Black	-0.02	0.157	-0.15	0.881	
AIAN_NHPI	0.13	0.268	0.50	0.620	
Asian	0.32	0.141	2.30	0.021	*
White	0.18	0.126	1.43	0.154	
Two or more races	0.17	0.175	1.00	0.319	
Cohort (reference = 2013-2014 kindergarten cohort)					
2014-2015 kindergarten cohort	0.30	0.032	9.19	<.001	***
2015-2016 kindergarten cohort	0.53	0.041	12.98	<.001	***
Age (in years; centered)	0.05	0.040	1.24	0.214	
Primary disability status (reference = no disabilities)					
Specific learning disabilities	-2.53	0.167	-15.12	<.001	***
Speech/language impairments	-0.61	0.058	-10.40	<.001	***
Other disabilities	-1.39	0.131	-10.66	<.001	***
Primary home language (reference = Spanish)					
AfroAsiatic	-0.26	0.326	-0.80	0.424	
Albanian	0.60	0.298	2.02	0.043	*
Arabic	0.41	0.178	2.28	0.023	*
Austronesian	0.64	0.269	2.38	0.017	*
BaltoSlavic	0.32	0.279	1.13	0.257	

Table 52 cont'd

Bengali	0.79	0.298	2.64	0.008	**
Chinese	1.06	0.275	3.85	<.001	***
Dravidian	1.41	0.247	5.69	<.001	***
IndoIranian	0.61	0.237	2.58	0.010	**
Italic/Germanic	0.87	0.265	3.29	<.001	***
Japanese/Korean	1.26	0.286	4.42	<.001	***
Other home languages	0.55	0.235	2.35	0.019	*
Instructional program					
(reference = ESL)					
Transitional bilingual	-0.34	0.168	-2.02	0.044	*
Dual-language bilingual	-0.43	0.356	-1.20	0.231	
No program	0.14	0.399	0.36	0.717	
Poverty status					
(reference = never in poverty)					
Retention	-0.51	0.124	-4.09	<.001	***
(reference = never retained)					
Home English use	2.37	0.185	12.82	<.001	***
(reference = never used					
English at home)					
Initial English proficiency	0.35	0.108	3.27	0.001	**
(centered)					
Interactions	0.81	0.038	21.36	<.001	***
AfroAsiatic*time	-0.01	0.081	-0.17	0.868	
Albanian*time	-0.14	0.080	-1.74	0.081	
Arabic*time	-0.15	0.036	-4.01	<.001	***
Austronesian*time	-0.05	0.072	-0.67	0.500	
BaltoSlavic*time	-0.06	0.076	-0.74	0.461	
Bengali*time	-0.13	0.078	-1.66	0.097	
Chinese*time	-0.11	0.081	-1.33	0.185	
Dravidian*time	-0.40	0.070	-5.70	<.001	***
IndoIranian*time	-0.17	0.060	-2.75	0.006	**
Italic/Germanic*time	-0.20	0.075	-2.62	0.009	**
Japanese/Korean*time	-0.23	0.089	-2.63	0.009	**
Other home					
languages*time	-0.15	0.059	-2.58	0.010	**
Transitional bilingual*time	0.07	0.044	1.66	0.096	
Dual-language					
bilingual*time	0.13	0.092	1.43	0.153	
No program*time	0.04	0.127	0.30	0.765	
Poverty status*time	0.03	0.038	0.73	0.466	
Retention*time	-0.79	0.053	-14.93	<.001	***

Table 52 cont'd

Home English use*time	-0.11	0.031	-3.51	<.001	***
Initial English proficiency (centered)*time	-0.08	0.011	-7.48	<.001	***
School district (reference = Dearborn City School)					
District3010	-14.87	1012.000	-0.02	0.988	
District3020	-14.61	1092.000	-0.01	0.989	
District3030	-14.34	1168.000	-0.01	0.990	
District3040	1.75	1.555	1.13	0.261	
District3050	-0.70	0.295	-2.38	0.017	*
District3070	-13.26	1264.000	-0.01	0.992	
District3080	-0.03	1.278	-0.03	0.979	
District3100	-0.52	0.460	-1.13	0.259	
District3900	-13.14	760.600	-0.02	0.986	
District4010	-1.03	1.174	-0.88	0.380	
District5060	-13.08	1064.000	-0.01	0.990	
District8030	1.40	0.529	2.64	0.008	**
District8050	-0.95	0.651	-1.46	0.143	
District9010	-14.18	525.900	-0.03	0.978	
District11010	-1.19	0.769	-1.55	0.121	
District11020	-0.44	0.460	-0.95	0.342	
District11030	-1.50	0.435	-3.45	<.001	***
District11033	-0.37	1.106	-0.34	0.736	
District11210	-2.03	1.060	-1.92	0.055	
District11240	0.04	0.249	0.18	0.857	
District11250	-2.90	1.021	-2.84	0.005	**
District11300	-1.04	0.391	-2.65	0.008	**
District11310	-10.88	2400.000	-0.01	0.996	
District11320	0.02	0.392	0.05	0.958	
District11330	-15.38	745.700	-0.02	0.984	
District11340	-12.20	2400.000	-0.01	0.996	
District11901	-0.88	0.514	-1.72	0.085	
District11903	-15.86	2400.000	-0.01	0.995	
District12010	-0.40	0.189	-2.12	0.034	*
District12020	-0.92	0.440	-2.10	0.036	*
District12901	-1.75	0.804	-2.17	0.030	*
District13020	-1.52	0.300	-5.06	<.001	***
District13070	-1.84	0.871	-2.11	0.035	*
District13090	-0.24	0.204	-1.16	0.247	
District13120	-15.04	1656.000	-0.01	0.993	



Table 52 cont'd

District13901	-1.35	0.813	-1.66	0.097	
District13902	0.08	0.425	0.19	0.851	
District13903	17.29	2400.000	0.01	0.994	
District14020	-0.38	0.375	-1.01	0.313	
District14030	0.45	1.218	0.37	0.715	
District14050	16.36	2400.000	0.01	0.995	
District17010	-11.52	1696.000	-0.01	0.995	
District19010	-1.15	1.186	-0.97	0.334	
District19070	1.19	0.736	1.61	0.107	
District19120	0.19	0.870	0.21	0.832	
District19125	-1.32	1.036	-1.27	0.203	
District19140	-1.38	1.078	-1.28	0.201	
District23030	3.08	4.356	0.71	0.480	
District23060	-0.88	0.365	-2.41	0.016	*
District24040	-15.97	1034.000	-0.02	0.988	
District24070	-12.90	1637.000	-0.01	0.994	
District25010	-10.85	1212.000	-0.01	0.993	
District25030	-0.46	0.207	-2.22	0.026	*
District25040	-15.11	796.900	-0.02	0.985	
District25050	-0.53	1.389	-0.38	0.705	
District25060	-14.99	1649.000	-0.01	0.993	
District25080	-0.33	0.468	-0.70	0.484	
District25100	-1.46	0.673	-2.17	0.030	*
District25110	-1.32	1.115	-1.18	0.237	
District25120	-0.36	0.450	-0.81	0.419	
District25130	-2.10	1.087	-1.94	0.053	
District25140	0.04	0.721	0.06	0.956	
District25150	-0.40	0.860	-0.47	0.639	
District25180	-1.06	1.059	-1.00	0.319	
District25200	-2.16	1.064	-2.03	0.043	*
District25230	0.62	1.488	0.42	0.677	
District25240	-13.48	964.100	-0.01	0.989	
District25250	0.51	0.865	0.59	0.553	
District25902	0.21	1.352	0.16	0.874	
District25903	-13.28	613.400	-0.02	0.983	
District25905	-0.54	1.388	-0.39	0.698	
District25909	-12.70	1264.000	-0.01	0.992	
District25910	-0.20	0.595	-0.34	0.732	
District28000	-10.50	2400.000	0.00	0.997	

Table 52 cont'd

District28010	-0.92	0.273	-3.38	<.001	***
District28900	-17.03	2400.000	-0.01	0.994	
District28902	-0.66	1.159	-0.57	0.571	
District29010	-15.22	590.000	-0.03	0.979	
District29040	-12.59	583.100	-0.02	0.983	
District29050	-15.70	622.700	-0.03	0.980	
District29060	-13.88	1489.000	-0.01	0.993	
District29100	-13.85	731.800	-0.02	0.985	
District30020	-12.56	1599.000	-0.01	0.994	
District32080	-11.91	1591.000	-0.01	0.994	
District33010	-0.25	0.214	-1.16	0.245	
District33020	-0.89	0.175	-5.08	<.001	***
District33040	17.44	2400.000	0.01	0.994	
District33060	-0.06	0.530	-0.11	0.910	
District33070	-1.35	0.303	-4.48	<.001	***
District33130	-1.59	1.079	-1.47	0.141	
District33170	-0.82	0.341	-2.41	0.016	*
District33200	-11.77	2400.000	-0.01	0.996	
District33215	0.53	0.343	1.53	0.125	
District33230	-1.63	1.143	-1.43	0.154	
District33904	-14.93	735.200	-0.02	0.984	
District33909	-1.06	0.426	-2.49	0.013	*
District33910	-1.41	0.512	-2.76	0.006	**
District33911	0.01	0.978	0.01	0.992	
District34010	-0.41	0.373	-1.09	0.274	
District34080	-2.15	0.629	-3.43	<.001	***
District34090	-0.65	0.876	-0.74	0.458	
District34110	-13.87	1507.000	-0.01	0.993	
District34120	-0.81	1.216	-0.66	0.507	
District37010	-0.90	0.701	-1.28	0.201	
District37901	-0.33	1.140	-0.29	0.769	
District38010	-0.32	1.274	-0.25	0.800	
District38020	-15.53	906.800	-0.02	0.986	
District38050	-13.98	781.200	-0.02	0.986	
District38090	-12.52	652.400	-0.02	0.985	
District38120	-16.07	1254.000	-0.01	0.990	
District38140	-14.89	428.700	-0.04	0.972	
District38170	-0.98	0.371	-2.65	0.008	**
District38902	-0.18	0.866	-0.21	0.832	

Table 52 cont'd

District39010	-0.55	0.155	-3.54	<.001	***
District39020	-16.17	1462.000	-0.01	0.991	
District39030	0.19	0.642	0.30	0.766	
District39065	-14.46	610.400	-0.02	0.981	
District39130	-0.97	1.294	-0.75	0.453	
District39140	-0.36	0.175	-2.09	0.037	*
District39170	-1.00	1.251	-0.80	0.422	
District39905	-13.68	590.400	-0.02	0.982	
District39908	20.16	2400.000	0.01	0.993	
District41010	-0.84	0.103	-8.20	<.001	***
District41020	0.09	0.151	0.61	0.543	
District41025	2.06	0.665	3.10	0.002	**
District41026	-0.71	0.161	-4.41	<.001	***
District41040	2.19	0.371	5.92	<.001	***
District41050	-0.41	0.378	-1.09	0.274	
District41070	-0.24	0.564	-0.42	0.676	
District41080	-0.07	0.258	-0.26	0.793	
District41090	-0.32	0.633	-0.51	0.610	
District41110	-0.39	0.162	-2.43	0.015	*
District41120	-1.19	0.180	-6.62	<.001	***
District41130	-1.58	0.475	-3.32	<.001	***
District41140	-0.81	0.186	-4.38	<.001	***
District41145	-0.94	0.415	-2.26	0.024	*
District41150	0.55	0.259	2.12	0.034	*
District41160	-0.24	0.112	-2.13	0.033	*
District41170	-0.49	0.466	-1.06	0.289	
District41210	-12.65	2400.000	-0.01	0.996	
District41240	-0.17	0.297	-0.56	0.575	
District41901	-0.12	0.437	-0.28	0.779	
District41904	-0.92	0.496	-1.86	0.064	
District41905	-0.22	0.227	-0.99	0.322	
District41908	1.24	1.191	1.04	0.299	
District41909	-0.72	0.194	-3.73	<.001	***
District41910	-1.19	0.328	-3.61	<.001	***
District41914	-1.22	0.510	-2.39	0.017	*
District41915	-0.16	0.374	-0.44	0.661	
District41916	-0.63	0.344	-1.85	0.065	
District41918	0.84	1.388	0.60	0.547	
District41919	-0.27	0.306	-0.89	0.374	

Table 52 cont'd

District41920	0.81	0.828	0.98	0.326	
District41921	-0.83	0.627	-1.33	0.184	
District41925	-1.47	1.135	-1.29	0.197	
District41926	-1.63	0.485	-3.36	<.001	***
District41927	-13.65	2400.000	-0.01	0.995	
District41928	-0.38	0.733	-0.52	0.607	
District44010	-11.29	1105.000	-0.01	0.992	
District44020	0.05	0.555	0.10	0.924	
District44060	-0.14	0.243	-0.58	0.565	
District45010	-12.98	1286.000	-0.01	0.992	
District45020	-2.08	1.057	-1.97	0.049	*
District45040	-14.28	1038.000	-0.01	0.989	
District45050	-1.66	1.066	-1.55	0.120	
District46010	-1.11	0.384	-2.89	0.004	**
District46080	-15.10	1299.000	-0.01	0.991	
District46090	-11.66	2400.000	-0.01	0.996	
District46110	-0.51	0.877	-0.58	0.564	
District47010	-1.49	0.493	-3.02	0.003	**
District47030	-11.28	1641.000	-0.01	0.995	
District47060	0.10	0.889	0.12	0.906	
District47070	0.02	0.363	0.06	0.956	
District47080	-14.88	747.600	-0.02	0.984	
District47902	-16.17	1574.000	-0.01	0.992	
District50010	-0.21	0.186	-1.15	0.249	
District50020	-0.95	0.783	-1.22	0.224	
District50030	-0.24	0.373	-0.65	0.519	
District50040	-0.78	0.458	-1.71	0.088	
District50050	-0.39	1.148	-0.34	0.737	
District50070	-1.19	0.685	-1.74	0.081	
District50080	-0.55	0.115	-4.81	<.001	***
District50090	-1.33	0.270	-4.91	<.001	***
District50100	-0.89	0.294	-3.04	0.002	**
District50120	-0.88	0.590	-1.49	0.135	
District50130	-0.84	0.721	-1.17	0.243	
District50140	-0.53	0.213	-2.48	0.013	*
District50160	-14.26	350.400	-0.04	0.968	
District50170	-1.98	0.648	-3.06	0.002	**
District50180	-1.21	0.673	-1.79	0.073	
District50190	-0.33	0.221	-1.49	0.137	

Table 52 cont'd

District50200	1.97	0.883	2.23	0.026	*
District50210	-0.31	0.084	-3.70	<.001	***
District50220	-0.94	0.640	-1.47	0.142	
District50230	-0.13	0.077	-1.62	0.105	
District50240	-2.33	0.361	-6.46	<.001	***
District50903	0.08	0.247	0.31	0.758	
District50908	1.17	0.856	1.37	0.171	
District50909	-13.14	1039.000	-0.01	0.990	
District50912	-0.65	1.123	-0.58	0.563	
District50913	0.29	0.251	1.15	0.252	
District51020	-12.95	798.200	-0.02	0.987	
District51045	-9.90	2400.000	0.00	0.997	
District51905	-14.42	1297.000	-0.01	0.991	
District52170	-11.51	2400.000	-0.01	0.996	
District53010	0.50	0.814	0.62	0.536	
District53040	-13.99	1666.000	-0.01	0.993	
District54010	-12.50	1320.000	-0.01	0.992	
District55120	-14.25	1012.000	-0.01	0.989	
District56010	-0.06	0.402	-0.15	0.883	
District57020	-15.36	789.900	-0.02	0.984	
District57030	-0.56	0.623	-0.91	0.365	
District58010	-2.12	0.487	-4.36	<.001	***
District58020	-1.26	0.845	-1.49	0.137	
District58030	-2.10	1.062	-1.98	0.048	*
District58080	-13.85	2400.000	-0.01	0.995	
District58902	0.93	0.616	1.50	0.133	
District59020	-1.00	1.143	-0.88	0.381	
District59070	-0.74	0.475	-1.56	0.119	
District59125	-0.96	1.101	-0.88	0.381	
District61010	-0.31	0.297	-1.03	0.304	
District61060	-3.38	1.035	-3.27	0.001	**
District61065	-0.26	0.843	-0.31	0.755	
District61080	-16.41	943.200	-0.02	0.986	
District61120	-12.43	2400.000	-0.01	0.996	
District61180	0.20	1.183	0.17	0.865	
District61190	-14.42	1167.000	-0.01	0.990	
District61210	-1.24	1.078	-1.15	0.251	
District61220	-14.84	931.300	-0.02	0.987	
District61240	0.54	0.963	0.57	0.572	

Table 52 cont'd

District61902	-1.10	0.445	-2.48	0.013	*
District62050	-0.87	0.328	-2.64	0.008	**
District62070	-0.64	0.578	-1.11	0.266	
District63010	-0.34	0.192	-1.79	0.074	
District63020	-0.84	0.806	-1.05	0.295	
District63030	-0.22	0.172	-1.30	0.195	
District63040	0.32	0.296	1.10	0.274	
District63050	-0.93	0.397	-2.36	0.018	*
District63060	-0.45	0.320	-1.40	0.163	
District63070	-0.14	0.201	-0.71	0.476	
District63080	-0.45	0.199	-2.25	0.025	*
District63090	0.97	0.265	3.67	<.001	***
District63100	0.13	0.123	1.09	0.274	
District63110	0.07	0.337	0.19	0.846	
District63130	-1.64	0.681	-2.40	0.016	*
District63140	-0.76	1.224	-0.62	0.534	
District63150	-0.09	0.091	-0.95	0.342	
District63160	0.04	0.145	0.27	0.791	
District63180	-1.34	0.778	-1.72	0.085	
District63190	-0.94	0.314	-3.00	0.003	**
District63200	-0.30	0.111	-2.73	0.006	**
District63210	0.04	0.496	0.09	0.930	
District63220	-0.43	0.232	-1.85	0.064	
District63230	-0.41	0.177	-2.32	0.020	*
District63240	-0.01	0.182	-0.05	0.959	
District63250	-0.26	0.563	-0.45	0.650	
District63260	0.09	0.114	0.79	0.430	
District63270	-0.25	0.345	-0.73	0.468	
District63280	-0.20	0.194	-1.04	0.298	
District63290	-0.57	0.113	-5.05	<.001	***
District63300	-0.64	0.162	-3.96	<.001	***
District63900	-10.22	2400.000	0.00	0.997	
District63901	-0.29	0.491	-0.60	0.550	
District63906	-2.40	1.034	-2.32	0.020	*
District63909	-1.27	0.329	-3.87	<.001	***
District63912	-0.41	0.218	-1.88	0.060	
District63913	-1.31	0.251	-5.23	<.001	***
District63914	-1.51	0.554	-2.72	0.007	**
District63915	-13.96	375.600	-0.04	0.970	

Table 52 cont'd

District63916	-11.65	1272.000	-0.01	0.993	
District63922	0.29	0.501	0.57	0.568	
District63923	-12.99	1283.000	-0.01	0.992	
District63926	-13.06	2400.000	-0.01	0.996	
District63927	-12.86	572.200	-0.02	0.982	
District63929	-0.60	0.692	-0.86	0.390	
District63933	-12.82	462.200	-0.03	0.978	
District63934	0.16	1.177	0.14	0.891	
District63938	-1.44	0.544	-2.64	0.008	**
District63939	-17.05	1340.000	-0.01	0.990	
District64040	-1.20	0.354	-3.39	<.001	***
District64080	-1.24	0.337	-3.67	<.001	***
District64090	-15.19	503.600	-0.03	0.976	
District70010	-0.38	0.348	-1.09	0.275	
District70020	-0.30	0.212	-1.41	0.159	
District70040	0.14	0.309	0.45	0.657	
District70070	0.13	0.121	1.04	0.297	
District70120	-1.20	0.390	-3.09	0.002	**
District70175	-0.27	0.887	-0.30	0.764	
District70190	0.45	0.397	1.14	0.254	
District70350	-1.06	0.274	-3.88	<.001	***
District70901	-0.44	1.424	-0.31	0.758	
District70902	-15.01	794.900	-0.02	0.985	
District70904	-0.76	0.923	-0.83	0.410	
District70905	-0.18	0.328	-0.56	0.574	
District70906	-0.58	0.248	-2.36	0.018	*
District70908	1.28	1.057	1.21	0.225	
District70909	15.72	2400.000	0.01	0.995	
District73010	-2.09	0.437	-4.78	<.001	***
District73030	-16.05	2400.000	-0.01	0.995	
District73040	-0.67	0.293	-2.31	0.021	*
District73190	-0.56	1.200	-0.47	0.641	
District73200	-16.07	917.600	-0.02	0.986	
District73910	-13.55	744.700	-0.02	0.985	
District74010	0.20	0.563	0.35	0.729	
District74040	-2.00	1.058	-1.89	0.059	
District74050	-14.81	656.800	-0.02	0.982	
District74100	-14.04	1300.000	-0.01	0.991	
District74903	-13.13	2400.000	-0.01	0.996	

Table 52 cont'd

District75010	-0.77	0.156	-4.91	<.001	***
District75030	-16.12	1675.000	-0.01	0.992	
District75040	-1.13	0.460	-2.45	0.014	*
District75050	-14.49	2400.000	-0.01	0.995	
District75070	-16.44	1304.000	-0.01	0.990	
District75080	-0.36	0.490	-0.73	0.467	
District75100	-0.36	0.358	-1.01	0.311	
District78020	-9.51	2400.000	0.00	0.997	
District78030	-15.13	2400.000	-0.01	0.995	
District78080	-16.23	2400.000	-0.01	0.995	
District80010	-0.74	0.355	-2.10	0.036	*
District80020	-1.48	0.365	-4.06	<.001	***
District80040	-1.81	0.536	-3.38	<.001	***
District80050	-0.04	0.466	-0.09	0.925	
District80090	-0.99	0.402	-2.47	0.013	*
District80110	-15.55	644.500	-0.02	0.981	
District80120	-0.37	0.248	-1.48	0.138	
District80130	-1.69	0.755	-2.23	0.026	*
District80140	-2.36	0.757	-3.12	0.002	**
District80150	0.14	0.509	0.28	0.779	
District80160	-0.40	0.613	-0.66	0.513	
District80240	-12.14	1643.000	-0.01	0.994	
District81010	-0.33	0.106	-3.16	0.002	**
District81020	-0.64	0.303	-2.12	0.034	*
District81040	0.77	0.524	1.48	0.140	
District81050	-15.96	1641.000	-0.01	0.992	
District81070	-0.27	0.658	-0.41	0.680	
District81080	-12.65	1264.000	-0.01	0.992	
District81100	-15.35	1265.000	-0.01	0.990	
District81120	-0.50	0.291	-1.72	0.085	
District81900	-12.60	668.900	-0.02	0.985	
District81901	0.70	1.535	0.46	0.647	
District81902	0.35	0.212	1.66	0.097	
District81905	-0.36	0.240	-1.52	0.128	
District81906	0.27	0.285	0.94	0.349	
District81908	-1.98	0.750	-2.64	0.008	**
District81910	0.28	0.366	0.77	0.439	
District81912	0.34	0.334	1.01	0.312	
District82015	-1.47	0.108	-13.59	<.001	***



Table 52 cont'd

District82020	-0.56	0.410	-1.37	0.170	
District82040	-1.41	0.275	-5.14	<.001	***
District82045	-0.35	0.137	-2.58	0.010	**
District82050	-0.12	0.366	-0.32	0.747	
District82055	-0.08	0.388	-0.21	0.836	
District82060	-0.86	0.148	-5.85	<.001	***
District82090	-0.50	0.134	-3.72	<.001	***
District82095	-0.02	0.134	-0.13	0.896	
District82100	-0.14	0.104	-1.36	0.175	
District82110	-0.88	0.548	-1.60	0.110	
District82120	-0.86	0.582	-1.48	0.140	
District82130	-0.30	0.630	-0.47	0.638	
District82140	-0.93	0.465	-2.00	0.046	*
District82150	-2.05	0.409	-5.02	<.001	***
District82155	-1.39	0.792	-1.75	0.080	
District82160	-1.12	0.200	-5.60	<.001	***
District82170	-1.18	0.431	-2.75	0.006	**
District82180	0.23	0.717	0.32	0.748	
District82230	0.09	0.103	0.82	0.410	
District82240	-1.13	1.111	-1.02	0.309	
District82250	0.14	1.137	0.13	0.900	
District82290	-0.33	0.488	-0.68	0.499	
District82300	-14.32	572.800	-0.03	0.980	
District82340	-0.21	0.606	-0.35	0.725	
District82365	-0.26	0.221	-1.16	0.246	
District82390	0.14	0.132	1.06	0.290	
District82400	-1.06	0.644	-1.65	0.099	
District82405	-0.49	0.325	-1.51	0.130	
District82430	-0.69	0.389	-1.77	0.077	
District82702	-13.79	1672.000	-0.01	0.993	
District82705	-15.27	2400.000	-0.01	0.995	
District82707	-9.53	849.300	-0.01	0.991	
District82717	-0.10	0.165	-0.60	0.548	
District82718	-0.92	0.543	-1.69	0.090	
District82725	-1.20	0.466	-2.57	0.010	*
District82729	-0.11	0.203	-0.54	0.593	
District82732	-3.30	1.024	-3.23	0.001	**
District82738	-0.28	1.032	-0.27	0.787	
District82743	0.04	0.196	0.21	0.834	

Table 52 cont'd

District82744	-0.13	0.270	-0.48	0.632	
District82745	0.35	0.339	1.03	0.303	
District82754	-16.09	802.700	-0.02	0.984	
District82755	-1.84	0.539	-3.42	<.001	***
District82757	-0.15	0.253	-0.59	0.558	
District82915	18.17	2400.000	0.01	0.994	
District82916	1.34	0.982	1.36	0.173	
District82918	-0.34	0.135	-2.52	0.012	*
District82921	-1.16	0.496	-2.33	0.020	*
District82928	-0.54	0.261	-2.09	0.037	*
District82938	-0.18	0.368	-0.50	0.620	
District82940	-1.15	1.108	-1.04	0.301	
District82941	-0.76	0.145	-5.26	<.001	***
District82950	-1.82	0.355	-5.13	<.001	***
District82956	-10.38	2400.000	0.00	0.997	
District82957	-0.69	0.190	-3.65	<.001	***
District82967	-14.59	492.600	-0.03	0.976	
District82968	-0.55	0.209	-2.65	0.008	**
District82969	-0.77	0.437	-1.77	0.076	
District82970	-0.78	0.837	-0.93	0.351	
District82973	-1.30	1.087	-1.20	0.231	
District82975	-0.16	0.197	-0.83	0.405	
District82976	0.36	0.422	0.85	0.396	
District82977	-0.01	0.184	-0.04	0.966	
District82981	-2.49	1.083	-2.30	0.021	*
District82982	-0.87	0.237	-3.67	<.001	***
District82983	-0.76	0.194	-3.93	<.001	***
District82986	-0.26	0.236	-1.12	0.262	
District82987	-1.88	1.123	-1.67	0.095	
District82995	0.02	0.686	0.04	0.971	
District83010	-14.86	715.600	-0.02	0.983	
District83060	0.80	1.202	0.66	0.507	
District83070	-11.49	1683.000	-0.01	0.995	
District83900	-0.79	1.209	-0.66	0.512	
District84060	0.25	1.130	0.23	0.822	

Note. AIAN/NHPI = American Indian or Alaska Native/Native Hawaiian or Other Pacific Islander; p value: <.001\*\*\*, .001 - .01\*\*, .01 - .05\*

## APPENDIX I. Graphing Method Description

In this section, I describe the method I used to visualize the results of the conditional discrete-time hazard model. Researchers using survival analysis often need to visualize the effects of particular predictors by graphing the fitted survival function. There are in general two methods to do so. The first method is called the mean of covariates method, and the second method is called the corrected group prognosis method (Ghali et al., 2001). With the first method, mean values for covariates are inserted into the survival function of the fitted hazard model. Despite its popularity, this method has been criticized for two major limitations. First, this method gives the survival curve of an average subject that may represent nobody in the dataset. For example, for a sample that consists of 46 females and 54 males, if males are chosen as the reference group and coded 0 and females chosen as the comparison group and coded 1, an average person would have a value of 0.46 on the variable sex. This average value, however, represents nobody in the original dataset. Second, the mean of covariates method provides the survival curve of an average subject rather than the average survival curve of all subjects. This is because averaging occurs before model predictions are made. Given that the hazard model is a nonlinear, exponential function, averaging a predictor's values before performing the exponential transformation is not the same as taking the averaging after exponentially transforming the predictor's values (see more in Ghali et al., 2001). The second method addresses these two limitations by showing the average treatment effect. With the second method, averaging occurs after the exponential transformation. This means that the fitted survival curve is obtained by averaging the survival curves of individual subjects. Specifically, I implemented this method in following the steps:

- 1) Create a new dataset (Data1) that contains all model variables in the original dataset except the predictor of interest (Predictor A).
- 2) Expand Data1 so that everyone in the dataset has data for the maximum number of years. For time-varying variables, a subject's the most recent (or latest) observation was carried forward to impute the missing data points.
- 3) Create another dataset (Data2) that contains as many copies of Data1 as the number of levels in Predictor A (e.g., given sex has two levels, female and male, Data2 would contain two copies Data1).
- 4) Add the variable Predictor A to Data2. Assign values to this variable so that each copy of Data1 get a different predictor level.
- 5) Use Data2 to do model prediction.
- 6) Calculate the predicted survival function for each individual in the Data2, then take the average of these survival curves to generate mean survival probabilities for each time period. Repeat this procedure for each level of Predictor A.

## APPENDIX J. Results for Sensitivity Analyses

As described in Chapter 7, the following model was used to answer research question 2.

$$\begin{aligned} \text{Logit } h(t_{ij}) = & [\alpha_1 D_{i1} + \alpha_2 D_{i2} + \dots + \alpha_j D_{ij}] \\ & + \beta_1 X_i + \beta_2 \text{Disability}_i + \beta_3 \text{Language}_i + \beta_4 \text{Program}_i \\ & + \beta_5 \text{Poverty}_{ij} + \beta_6 \text{HomeEnglish}_{ij} + \beta_7 \text{Retention}_{ij} \\ & + \beta_8 \text{Language}_i * T_{ij} + \beta_9 \text{Program}_i * T_{ij} + \beta_{10} \text{Poverty}_{ij} * T_{ij} \\ & + \beta_{11} \text{HomeEnglish}_{ij} * T_{ij} + \beta_{12} \text{Retention}_{ij} * T_{ij} \\ & + \text{Cohort}_i + \text{SchoolDistrict}_{ij} \\ & (+ \beta_{13} \text{InitialProficiency}_i + \beta_{14} \text{InitialProficiency}_i * T_{ij}) \end{aligned} \quad (\text{Model 1})$$

I conducted three sensitivity analyses to test if removing school district or cohort from Model 1 would result in any changes in the effects of the six core predictors (primary disability type, primary home language, instructional programming, poverty status, home English use, and retention) on the likelihood of proficiency attainment. In the first analysis, I ran the model without school district (Model 2a); in the second analysis, I ran the model without cohort (Model 2b); and in the third analysis, I ran the model without either of these two variables (Model 2c). All three analyses were based on the data of the subsample.

Table 53 provides a side-by-side comparison of the predictor parameters estimated from Models 1, 2a, 2b, and 2c. One can see that removing school district, cohort, or both control variables from the model did not lead to meaningful changes in the magnitude, direction, or significance of parameter estimates for any of the predictors nor their interactions with time. (I consider a predictor's coefficients to be "different" between two models if the absolute value of the difference is at least twice as large as the predictor's standard error.) The only changes that were notable were in the magnitudes of the estimates of transitional bilingual program and its interaction with time. In general, the effects of transitional bilingual program (as compared to the reference category of ESL program) and its interaction with time on the likelihood of proficiency attainment were slightly stronger (as evidenced by larger absolute values) in models that did not control for school district (Model 2a), cohort (Model 2b), or either variable (Model 2c) than in the model that controlled for these two variables (Model 1). This is perhaps because students who ever participated in a transitional bilingual program were largely clustered in one cohort (i.e., the cohort of 2013-2014 kindergarten entrants) within one district (i.e., Dearborn School Districts). Please see Chapter 7 for more information about the distribution of students who ever participated in a transitional bilingual program across cohorts and school districts. Given that Model 1 controlled for both school district and cohort, the effects of traditional bilingual program and its interaction with time were estimated within rather than between school districts and cohorts. The fact that school district and cohort explained part of the variance in the likelihood of proficiency associated with instructional programming may have led to smaller effect estimates (in absolute value) for these two predictors in Model 1 as compared to those in Models 2a, 2b, and 2c. Despite the differences in the magnitude of the effects of instructional programming across models, the general story holds: students in ESL programs were initially more likely to attain proficiency than those in bilingual programs in early elementary grades, but students in bilingual programs caught up and closed the gaps over time.

Table 53. Results for three sensitivity analyses in comparison with results of the model used to answer research question 2 (based on subsample data)

	Model 1			Model 2a		Model 2b		Model 2c	
	Logit	SE	p value	Logit	p value	Logit	p value	Logit	p value
Primary disability type (reference = no disabilities)									
Specific learning disabilities	-2.531	0.167	<.001***	-2.404	<.001***	-2.549	<.001***	-2.424	<.001***
Speech/language impairments	-0.608	0.058	<.001***	-0.550	<.001***	-0.604	<.001***	-0.546	<.001***
Other disabilities	-1.392	0.131	<.001***	-1.308	<.001***	-1.392	<.001***	-1.308	<.001***
Primary home language (reference = Spanish)									
AfroAsiatic	-0.261	0.326	0.424	-0.212	0.505	-0.250	0.443	-0.241	0.450
Albanian	0.603	0.298	0.043*	0.743	0.011*	0.639	0.032*	0.765	0.009**
Arabic	0.407	0.178	0.023*	0.677	<.001***	0.450	0.012*	0.729	<.001***
Austronesian	0.641	0.269	0.017*	0.832	0.001**	0.656	0.015*	0.843	0.001**
BaltoSlavic	0.316	0.279	0.257	0.444	0.103	0.365	0.191	0.476	0.081
Bengali	0.786	0.298	0.008**	0.924	<.001***	0.844	0.004**	0.973	<.001***
Chinese	1.061	0.275	<.001***	1.349	<.001***	1.091	<.001***	1.355	<.001***
Dravidian	1.407	0.247	<.001***	1.685	<.001***	1.491	<.001***	1.756	<.001***
IndoIranian	0.612	0.237	0.010**	0.847	<.001***	0.657	0.006**	0.878	<.001***
Italic/Germanic	0.872	0.265	<.001***	1.143	<.001***	0.908	<.001***	1.152	<.001***
Japanese/Korean	1.261	0.286	<.001***	1.610	<.001***	1.309	<.001***	1.642	<.001***
Other home languages	0.552	0.235	0.019*	0.783	<.001***	0.585	0.013*	0.812	<.001***
Instructional program (reference = ESL)									
Transitional bilingual	-0.339	0.168	0.044*	-0.654	<.001***	-0.763	<.001***	-0.931	<.001***
Dual-language bilingual	-0.426	0.356	0.231	-0.612	0.075	-0.483	0.173	-0.682	0.047*
No program	0.145	0.399	0.717	0.135	0.732	0.109	0.784	0.091	0.817

Table 53 cont'd

Poverty status (reference = never in poverty)	-0.508	0.124	<.001***	-0.508	<.001***	-0.456	<.001***	-0.457	<.001***
Retention (reference = never retained)	2.365	0.185	<.001***	2.386	<.001***	2.411	<.001***	2.421	<.001***
Home English use (reference = never used English at home)	0.354	0.108	0.001**	0.414	<.001***	0.431	<.001***	0.476	<.001***
Interactions									
AfroAsiatic*time	-0.013	0.081	0.868	0.047	0.554	-0.008	0.923	0.057	0.475
Albanian*time	-0.140	0.080	0.081	-0.102	0.198	-0.142	0.076	-0.101	0.204
Arabic*time	-0.146	0.036	<.001***	-0.104	0.003**	-0.148	<.001***	-0.109	0.002**
Austronesian* time	-0.049	0.072	0.500	-0.026	0.713	-0.053	0.463	-0.031	0.662
BaltoSlavic*time	-0.056	0.076	0.461	-0.041	0.588	-0.060	0.428	-0.040	0.594
Bengali*time	-0.130	0.078	0.097	-0.117	0.120	-0.142	0.068	-0.126	0.094
Chinese*time	-0.107	0.081	0.185	-0.107	0.169	-0.115	0.154	-0.113	0.149
Dravidian*time	-0.398	0.070	<.001***	-0.354	<.001***	-0.415	<.001***	-0.370	<.001***
IndoIranian*time	-0.166	0.060	0.005**	-0.133	0.024*	-0.171	0.004**	-0.138	0.019*
Italic/Germanic* time	-0.198	0.075	0.009**	-0.212	0.004**	-0.201	0.008**	-0.211	0.004**
Japanese/Korean* time	-0.235	0.089	0.009**	-0.212	0.015*	-0.239	0.007**	-0.219	0.011*
Other home languages*time	-0.152	0.059	0.010**	-0.158	0.006**	-0.156	0.008**	-0.161	0.005**
Transitional bilingual*time	0.072	0.044	0.096	0.145	<.001***	0.140	0.001**	0.196	<.001***
Dual-language bilingual*time	0.132	0.092	0.153	0.127	0.156	0.141	0.125	0.142	0.112
No program*time	0.038	0.127	0.765	0.038	0.760	0.040	0.752	0.044	0.721

Table 53 cont'd

Poverty status*time	0.028	0.038	0.466	-0.005	0.888	0.021	0.582	-0.011	0.773
Retention*time	-0.789	0.053	<.001***	-0.792	<.001***	-0.797	<.001***	-0.798	<.001***
Home English use*time	-0.110	0.031	<.001***	-0.117	<.001***	-0.125	<.001***	-0.131	<.001***

## APPENDIX K. Results for Research Question 3 Based on the Subsample

In this section, I present the results for research question 3 based on the subsample. Research question 3 asks about the time it takes ELs to meet the following individual proficiency criteria: (a) reaching a designated minimum composite score on ACCESS (2013-2019); (b) reaching designated cut scores on the reading and writing domains of ACCESS (2013-2019); (c) reaching designated cut scores on the speaking and listening domains of ACCESS (2013-2016). Table 54 displays the results of fitting the discrete-time hazard model (specified in Equation 1 in Chapter 7) to the data of the subsample to predict the hazard of reaching each of the five proficiency criteria as a function of time. Similar to the model results based on the full sample, results based on the subsample indicate that speaking and listening skills advanced more quickly than reading and writing skills. Half of students met the listening criterion within one year and the speaking criterion within two years. The median time to reading proficiency was slightly longer, but still within two years. Writing was the largest barrier to proficiency, with half of students requiring four to five years to develop proficiency in this domain.

Table 54. Results of the discrete-time hazard model (specified in Equation 1 in Chapter 7) predicting hazard of the subsample reaching individual proficiency criteria as a function of time, using data from 2013-2014 through 2018-2019

Criterion	Time	Years since school entry	Hazard	95% CI for hazard		Survival	Cumulative proportion attaining criterion
				Lower	Upper		
Listening	0	1	0.5846	0.5787	0.5905	0.4154	0.5846
	1	2	0.4219	0.4103	0.4335	0.2401	0.7599
	2	3	0.7103	0.6911	0.7288	0.0696	0.9304
Speaking	0	1	0.3262	0.3206	0.3318	0.6738	0.3262
	1	2	0.3727	0.3638	0.3817	0.4227	0.5773
	2	3	0.4882	0.4716	0.5047	0.2164	0.7836
Reading	0	1	0.2260	0.2210	0.2310	0.7740	0.2260
	1	2	0.4185	0.4115	0.4254	0.4501	0.5499
	2	3	0.3612	0.3521	0.3703	0.2876	0.7124
	3	4	0.2345	0.2244	0.2449	0.2201	0.7799
	4	5	0.3310	0.3144	0.3481	0.1473	0.8527
Writing	5	6	0.2414	0.2146	0.2704	0.1117	0.8883
	0	1	0.0190	0.0174	0.0207	0.9810	0.0190
	1	2	0.0166	0.0151	0.0183	0.9647	0.0353
	2	3	0.0872	0.0836	0.0910	0.8805	0.1195
	3	4	0.2426	0.2366	0.2486	0.6669	0.3331
	4	5	0.4222	0.4124	0.4321	0.3853	0.6147
Composite	5	6	0.4756	0.4578	0.4935	0.2021	0.7979
	0	1	0.0818	0.0786	0.0851	0.9182	0.0818



Table 54 cont'd

1	2	0.1082	0.1042	0.1123	0.8188	0.1812
2	3	0.1514	0.1465	0.1565	0.6948	0.3052
3	4	0.0957	0.0912	0.1004	0.6284	0.3716
4	5	0.3164	0.3068	0.3261	0.4296	0.5704
5	6	0.2914	0.2755	0.3078	0.3044	0.6956

## REFERENCES

## REFERENCES

- Abedi, J. (2008). Classification system for English language learners: Issues and recommendations. *Educational Measurement: Issues and Practice*, 27(3), 17–31.  
<https://doi.org/10.1111/j.1745-3992.2008.00125.x>
- Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher*, 33(1), 4–14.  
<https://doi.org/10.3102/0013189X033001004>
- Allen, C. S., Chen, Q., Willson, V. L., & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multilevel analysis. *Educational Evaluation and Policy Analysis*, 31(4), 480–499.  
<https://doi.org/10.3102/0162373709352239>
- Anderson, G. E., Jimerson, S. R., & Whipple, A. D. (2005). Student ratings of stressful experiences at home and school: Loss of a parent and grade retention as superlative stressors. *Journal of Applied School Psychology*, 21(1), 1–20.  
[https://doi.org/10.1300/J370v21n01\\_01](https://doi.org/10.1300/J370v21n01_01)
- Armstrong, R. A. (2019). Is there a large sample size problem? *Ophthalmic and Physiological Optics*, 39(3), 129–130. <https://doi.org/10.1111/opo.12618>
- August, D., & Hakuta, K. (1998). *Educating language minority children*. National Academy Press. <https://eric.ed.gov/?id=ED420737>
- August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on language minority children and youth*. Lawrence Erlbaum Associates.
- Bailey, A. L., & Butler, F. A. (2004). Ethical considerations in the assessment of the language and content knowledge of U.S. school-age English learners. *Language Assessment Quarterly*, 1(2–3), 177–193. <https://doi.org/10.1080/15434303.2004.9671784>
- Bailey, A. L., & Kelly, K. R. (2011). Home language survey practices in the initial identification of English learners in the United States. *Educational Policy*, 27(5), 770–804.  
<https://doi.org/10.1177/0895904811432137>
- Baker, M., & Johnston, P. (2010). The impact of socioeconomic status on high stakes testing reexamined. *Journal of Instructional Psychology*, 37(3), 193–199.
- Barnett, W. S., Yarosz, D. J., Thomas, J., Jung, K., & Blanco, D. (2007). Two-way and monolingual English immersion in preschool education: An experimental comparison.

- Early Childhood Research Quarterly*, 22(3), 277–293.  
<https://doi.org/10.1016/j.ecresq.2007.03.003>
- Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics*, 47(2), 383–411. <https://doi.org/10.1515/LING.2009.014>
- Bialik, K., Scheller, A., & Walker, K. (2018). *6 facts about English language learners in U.S. public schools*. Pew Research Center. <https://www.pewresearch.org/fact-tank/2018/10/25/6-facts-about-english-language-learners-in-u-s-public-schools/>
- Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology*, 65(4), 229–235. <https://doi.org/10.1037/a0025406>
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press.
- Blackledge, A., Creese, A., & Takhi, J. K. (2013). Language, superdiversity and education. In I. Saint-Georges & J. J. Weber (Eds.), *Multilingualism and multimodality* (pp. 59–80). Springer. [https://doi.org/10.1007/978-94-6209-266-2\\_4](https://doi.org/10.1007/978-94-6209-266-2_4)
- Bohman, T. M., Bedore, L. M., Peña, E. D., Mendez-Perez, A., & Gillam, R. B. (2010). What you hear and what you say: Language performance in Spanish-English bilinguals. *International Journal of Bilingual Education and Bilingualism*, 13(3), 325–344. <https://doi.org/10.1080/13670050903342019>
- Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective: A bioecological model. *Psychological Review*, 101(4), 568–586. <https://doi.org/10.1037/0033-295X.101.4.568>
- Bronfenbrenner, U., & Evans, G. W. (2000). Developmental science in the 21st century: Emerging questions, theoretical models, research designs and empirical findings. *Social Development*, 9(1), 115–125. <https://doi.org/10.1111/1467-9507.00114>
- Brooks, C. (2016). Summary of changes to Title III. <https://www.cde.state.co.us/fedprograms/ESSABlogPosts/titleiiichanges>
- Burke, A. M., Morita-Mullaney, T., & Singh, M. (2016). *Indiana emergent bilingual student time to reclassification: A survival analysis*. *American Educational Research Journal* (Vol. 53). <https://doi.org/10.3102/0002831216667481>
- Butler, Y. G. (2014). Parental factors and early English education as a foreign language: A case study in Mainland China. *Research Papers in Education*, 29(4), 410–437. <https://doi.org/10.1080/02671522.2013.776625>

- Butler, Y. G., & Le, V. N. (2018). A longitudinal investigation of parental social-economic status (SES) and young students' learning of English as a foreign language. *System*, 73, 4–15. <https://doi.org/10.1016/j.system.2017.07.005>
- Capps, R., Fix, M., Murray, J., Ost, J., Passel, J., & Herwantoro, S. (2005). *The new demography of America's schools: Immigration and the No Child Left Behind Act*. <https://files.eric.ed.gov/fulltext/ED490924.pdf>
- Center for Applied Linguistics. (2018). *Annual technical report for ACCESS for ELLs 2.0 online English language proficiency test, series 401, 2016-2017 administration*. <https://www.cde.state.co.us/assessment/accessforellsonlinetechreport>
- Cheung, A. C. K., & Slavin, R. E. (2012). Effective reading programs for Spanish-dominant English language learners (ELLs) in the elementary grades: A synthesis of research. *Review of Educational Research*, 82(4), 351–395. <https://doi.org/10.3102/0034654312465472>
- Chiswick, B. R., & Miller, P. W. (2005). Linguistic distance: A quantitative measure of the distance between English and other languages. *Journal of Multilingual and Multicultural Development*, 26(1), 1–11. <https://doi.org/10.1080/14790710508668395>
- Chubb, J., Linn, R., Haycock, K., & Wiener, R. (2005). Do we need to repair the monument? Debating the future of No Child Left Behind. *Education Next*, 5(2), 9–19. [https://link.gale.com/apps/doc/A130276020/OVIC?u=msu\\_main&sid=OVIC&xid=2fcf0dc1](https://link.gale.com/apps/doc/A130276020/OVIC?u=msu_main&sid=OVIC&xid=2fcf0dc1)
- Civil Rights Act of 1964, Pub. L. No. 88-352, 78 Stat. 241 (1964). <https://www.govinfo.gov/content/pkg/STATUTE-78/pdf/STATUTE-78-Pg241.pdf>
- Collier, V. P. (1987). Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly*, 21(4), 617–641. <https://doi.org/10.2307/3586986>
- Conger, D. (2009). Testing, time limits, and English learners: Does age of school entry affect how quickly students can learn English? *Social Science Research*, 38(2), 383–396. <https://doi.org/10.1016/j.ssresearch.2008.08.002>
- Conger, D. (2010). Does bilingual education interfere with English-language acquisition? *Social Science Quarterly*, 91(4), 1103–1122. <https://doi.org/10.1111/j.1540-6237.2010.00751.x>
- Conger, D., Hatch, M., McKinney, J., Atwell, M. S., & Lamb, A. (2008). Time to English proficiency for English language learners in New York City and Miami-Dade County. *IESP Policy Brief*, 1(12). <https://eric.ed.gov/?id=ED556750>
- Cook, G., Linquanti, R., Chinen, M., & Jung, H. (2012). *National evaluation of Title III implementation supplemental report—Exploring approaches to setting English language proficiency performance criteria and monitoring English learner progress*. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy

- and Program Studies Service. <https://www2.ed.gov/rschstat/eval/title-iii/implementation-supplemental-report.pdf>
- Cook, G., & MacDonald, R. (2014). *Reference performance level descriptors: Outcome of a national working session on defining an “English proficient” performance standard*. Council of Chief State School Officers. <https://www.wested.org/resources/toward-a-common-definition-of-english-learner-guidance-for-states-and-state-assessment-consortia-in-defining-and-addressing-policy-and-technical-issues-and-options/>
- Cook, G., & MacGregor, D. (n.d.). *ACCESS for ELLs 2.0 assessment proficiency level scores standard setting project report*. WIDA Research and the Center for Applied Linguistics. <https://www.cde.state.co.us/assessment/accessforellsstandardsettingreport>
- Crawford, J. (2007). The decline of bilingual education: How to reverse a troubling trend? *International Multilingual Research Journal*, 1(1), 33–37. <https://doi.org/10.1080/19313150709336863>
- Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reclassification. *Applied Linguistics*, 11(2), 132–149. <https://doi.org/10.1093/applin/IL2.132>
- Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, 49(2), 222–251. <https://doi.org/10.3102/00346543049002222>
- Cummins, J. (2019). Should schools undermine or sustain multilingualism? An analysis of theory, research, and pedagogical practice. *Sustainable Multilingualism*, 15(1), 1–26. <https://doi.org/10.2478/sm-2019-0011>
- Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. In B. Street & H. N. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed., Vol. 2, pp. 71–83). Springer Science and Business Media.
- De Cat, C. (2020). Predicting language proficiency in bilingual children. *Studies in Second Language Acquisition*, 42(2), 279–325. <https://doi.org/10.1017/S0272263119000597>
- Dixon, L. Q., & Wu, S. (2014). Home language and literacy practices among immigrant second-language learners. *Language Teaching*, 47(4), 414–449. <https://doi.org/10.1017/S0261444814000160>
- Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Blackwell.
- Duursma, E., Romero-Contreras, S., Szuber, A., Proctor, P., Snow, C., August, D., & Calderón, M. (2007). The role of home literacy and language environment on bilinguals’ English and

- Spanish vocabulary development. *Applied Psycholinguistics*, 28(1), 171–190.  
<https://doi.org/10.1017/S0142716407070099>
- Ekström, J. (2011). A generalized definition of the polychoric correlation coefficient.  
[https://escholarship.org/content/qt583610fv/qt583610fv\\_noSplash\\_1752d2d039fd66d190ef211cf8c14008.pdf?t=lrh0h9](https://escholarship.org/content/qt583610fv/qt583610fv_noSplash_1752d2d039fd66d190ef211cf8c14008.pdf?t=lrh0h9)
- Ellis, N. C., O'Donnell, M. B., & Römer, U. (2013). Usage-based language: Investigating the latent structures that underpin acquisition. *Language Learning*, 63(S1), 25–51.  
<https://doi.org/10.1111/j.1467-9922.2012.00736.x>
- Equal Educational Opportunities Act, 20 U.S.C. § 1701-1758 (1974).  
<https://www.govinfo.gov/content/pkg/USCODE-2010-title20/pdf/USCODE-2010-title20-chap39-subchapI-part2-sec1703.pdf>
- Estrada, P., & Wang, H. (2018). Making English learner reclassification to fluent English proficient attainable or elusive: When meeting criteria is and is not enough. *American Educational Research Journal*, 55(2), 207–242. <https://doi.org/10.3102/0002831217733543>
- Every Student Succeeds Act, 20 U.S.C. § 6301 (2015).  
<https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf>
- Figlio, D., & Özek, U. (2020). An extra year to learn English? Early grade retention and the human capital development of English learners. *Journal of Public Economics*, 186.  
<https://doi.org/10.1016/j.jpubeco.2020.104184>
- Flores, N., & Rosa, J. (2015). Undoing appropriateness: Raciolinguistic ideologies and language diversity in education. *Harvard Educational Review*, 85(2), 149–171.  
<https://doi.org/10.17763/0017-8055.85.2.149>
- Gándara, P., & Orfield, G. (2010). *A return to the “Mexican room”: The segregation of Arizona’s English learners*. Center for Latino Policy Research.  
<https://files.eric.ed.gov/fulltext/ED511322.pdf>
- Gathercole, V. C. M. (2002). Monolingual and bilingual acquisition: Learning different treatments of that-trace phenomena in English and Spanish. In P. K. Oller & D. R. Eiler (Eds.), *Language and literacy in bilingual children* (pp. 220–254). Multilingual Matters.  
<https://doi.org/10.21832/9781853595721-011>
- Ghali, W., Quan, H., Brant, R., van Melle, G., Norris, C. M., Faris, P. D., ... Knudtson, M. L. (2001). Comparison of 2 methods for calculating adjusted survival curves from proportional hazards models. *JAMA*, 286(12), 1494–1497. <https://doi.org/10.1001/jama.286.12.1494>
- Gillanders, C., Franco, X., Seidel, K., Castro, D. C., & Méndez, L. I. (2017). Young dual language learners’ emergent writing development. *Early Child Development and Care*, 187(3–4), 371–382. <https://doi.org/10.1080/03004430.2016.1211124>

- Goldenberg, B. C. (2008). Teaching English language learners: What the research does—and does not—say. *American Educator*.  
<https://www.aft.org/sites/default/files/periodicals/goldenberg.pdf>
- Goldenberg, C., Rueda, R. S., & August, D. (2006). Synthesis: Sociocultural contexts and literacy development. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the national literacy panel on language-minority children and youth* (pp. 249–267). Lawrence Erlbaum Associates.
- Gor, K., & Vatz, K. (2009). Less commonly taught languages: Issues in learning and teaching. In M. Long & C. Doughty (Eds.), *The handbook of language teaching* (pp. 234–249). Blackwell Publishing.
- Greenberg Motamedi, J. G., Singh, M., & Thompson, K. D. (2016). *English learner student characteristics and time to reclassification: An example from Washington state (REL 2016–128)*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northwest. <http://ies.ed.gov/ncee/edlabs>
- Greene, J. P., & Winters, M. A. (2007). Revisiting grade retention: An evaluation of Florida’s test-based promotion policy. *Education Finance and Policy*, 2(4), 319–340.  
<https://doi.org/10.1162/edfp.2007.2.4.319>
- Grimm, R. P., Solari, E. J., & Gerber, M. M. (2018). A longitudinal investigation of reading development from kindergarten to grade eight in a Spanish-speaking bilingual population. *Reading and Writing*, 31(3), 559–581. <https://doi.org/10.1007/s11145-017-9798-1>
- Gutiérrez-Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics*, 24(2), 267–288.  
<https://doi.org/10.1017/S0142716403000158>
- Haas, E. (2014). Proposition 227, Proposition 203, and Question 2 in the context of legal rights for English language learners. In G. P. McField (Ed.), *The Miseducation of English Learners: A Tale of Three States and Lessons to Be Learned*. (pp. 1–24). Information Age Publishing.
- Hakuta, K., Butler, Y. G., & Witt, D. (2000). How long does it take English learners to attain proficiency (Policy Report 2000-1). *The University of California Linguistic Minority Research Institute*. <https://eric.ed.gov/?ID=ED443275>
- Hammer, C. S., Davison, M. D., Lawrence, F. R., & Miccio, A. W. (2009). The effect of maternal language on bilingual children’s vocabulary and emergent literacy development during head start and kindergarten. *Scientific Studies of Reading*, 13(2), 99–121.  
<https://doi.org/10.1080/10888430902769541>



- Hammer, C. S., Hoff, E., Uchikoshi, Y., Gillanders, C., Castro, D. C., & Sandilos, L. E. (2014). The language and literacy development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, 29(4), 715–733. <https://doi.org/10.1016/j.ecresq.2014.05.008>
- Hammer, C. S., Jia, G., & Uchikoshi, Y. (2011). Language and literacy development of dual language learners growing up in the United States: A call for research. *Child Development Perspectives*, 5(1), 4–9. <https://doi.org/10.1111/j.1750-8606.2010.00140.x>
- Hammer, C. S., Komaroff, E., Rodriguez, B. L., Lopez, L. M., Scarpino, S. E., & Goldstein, B. (2012). Predicting Spanish–English bilingual children’s language abilities, 55(5), 1251–1264. [https://doi.org/10.1044/1092-4388\(2012/11-0016\)](https://doi.org/10.1044/1092-4388(2012/11-0016))
- Hoff, E. (2003). The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, 74(5), 1368–1378. <https://doi.org/10.1111/1467-8624.00612>
- Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Developmental Psychology*, 49(1), 4–14. <https://doi.org/10.1037/a0027238>
- Hoff, E. (2018). Bilingual development in children of immigrant families. *Child Development Perspectives*, 12(2), 80–86. <https://doi.org/10.1111/cdep.12262>
- Hussar, B., Zhang, J., Hein, S., Wang, K., Roberts, A., Cui, J., ... Purcell, S. (2020). *The condition of education 2020 (NCES 2020-144)*. National Center for Education Statistics. <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2020144>
- Hwang, J. K., Mancilla-Martinez, J., Flores, I., & McClain, J. B. (2020). The relationship among home language use, parental beliefs, and Spanish-speaking children’s vocabulary. *International Journal of Bilingual Education and Bilingualism*. <https://doi.org/10.1080/13670050.2020.1747389>
- Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics*, 86(1), 226–244. <https://doi.org/10.1162/003465304323023778>
- Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics*, 24(1), 131–161. <https://doi.org/10.1017/s0142716403000079>
- Jimerson, S. R., & Renshaw, T. L. (2012). Retention and social promotion. *Principal Leadership*, 13(1), 12–16. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.445.9362&rep=rep1&type=pdf>

- Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, 30(3), 420–437.  
<https://doi.org/10.1080/02796015.2001.12086124>
- Jimerson, S. R. (1999). On the failure of failure: Examining the association between early grade retention and education and employment outcomes during late adolescence. *Journal of School Psychology*, 37(3), 243–272. [https://doi.org/10.1016/S0022-4405\(99\)00005-9](https://doi.org/10.1016/S0022-4405(99)00005-9)
- Jimerson, S. R., & Ferguson, P. (2007). A longitudinal study of grade retention: Academic and behavioral outcomes of retained students through adolescence. *School Psychology Quarterly*, 22(3), 314–339. <https://doi.org/10.1037/1045-3830.22.3.314>
- Kangas, S. E. N. (2019). English learners with disabilities: Linguistic development and educational equity in jeopardy. In X. Gao (Ed.), *Second Handbook of English Language Teaching* (pp. 919–937). Springer. [https://doi.org/10.1007/978-3-030-02899-2\\_48](https://doi.org/10.1007/978-3-030-02899-2_48)
- Kangas, S. E. N. (2018). Breaking one law to uphold another: How schools provide services to English learners with disabilities. *TESOL Quarterly*, 52(4), 877–910.  
<https://doi.org/10.1002/tesq.431>
- Kangas, S. E. N. (2014). When special education trumps ESL: An investigation of service delivery for ELLs with disabilities. *Critical Inquiry in Language Studies*, 11(4), 273–306.  
<https://doi.org/10.1080/15427587.2014.968070>
- Kangas, S. E. N., & Cook, M. (2020). Academic tracking of English learners with disabilities in middle school. *American Educational Research Journal*, 57(6), 2415–2449.  
<https://doi.org/10.3102/0002831220915702>
- Kass, R., & Raftery, A. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430), 773–795.
- Kieffer, M. J., Lesaux, N. K., & Snow, C. E. (2007). Promises and pitfalls: Implications of No Child Left Behind for defining, assessing, and serving English language learners. In G. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, and school reform* (pp. 57–74). Corwin Press.
- Kieffer, M. J., & Parker, C. E. (2016). *Patterns of English Learner Student Reclassification in New York City Public Schools (REL 2017–200)*. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands.  
<https://files.eric.ed.gov/fulltext/ED569344.pdf>
- Kieffer, M. J. (2011). Converging trajectories: Reading growth in language minority learners and their classmates, kindergarten to grade 8. *American Educational Research Journal*, 48(5), 1187–1225. <https://doi.org/10.3102/0002831211419490>

- Kieffer, M. J. (2010). Socioeconomic status, English proficiency, and late-emerging reading difficulties. *Educational Researcher*, 39(6), 484–486. <https://doi.org/10.3102/0013189X10378400>
- Kieffer, M. J. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learners in the United States. *Journal of Educational Psychology*, 100(4), 851–868. <https://doi.org/10.1037/0022-0663.100.4.851>
- Kieffer, M. J. (2012). Early oral language and later reading development in Spanish-speaking English language learners: Evidence from a nine-year longitudinal study. *Journal of Applied Developmental Psychology*, 33(3), 146–157. <https://doi.org/10.1016/j.appdev.2012.02.003>
- Kim, H., Barron, C., Sinclair, J., & Jang, E. (2020). Change in home language environment and English literacy achievement over time: A multi-group latent growth curve modeling investigation. *Language Testing*, 37(4), 573–599. <https://doi.org/10.1177/0265532220930348>
- Kim, Y. K., Curby, T. W., & Winsler, A. (2014). Child, family, and school characteristics related to English proficiency development among low-income, dual language learners. *Developmental Psychology*, 50(12), 2600–2613. <https://doi.org/10.1037/a0038050>
- Klingner, J., Artiles, A., & Barletta, L. M. (2006). English language learners who struggle with reading: Language acquisition or LD? *Journal of Learning Disabilities*, 39(2), 108–128. <https://doi.org/10.1177/00222194060390020101>
- Lanauze, M., & Snow, C. (1989). The relation between first-and second-language writing skills: Evidence from Puerto Rican elementary school children in bilingual programs. *Linguistics and Education*, 1(4), 323–339. [https://doi.org/10.1016/S0898-5898\(89\)80005-1](https://doi.org/10.1016/S0898-5898(89)80005-1)
- Lau v. Nichols et al., 414 U.S. 563. 1974. <http://caselaw.lp.findlaw.com/cgi-bin/getcase.pl?court=US&vol=414&invol=563>
- Linquanti, R., & Cook, G. (2013). *Toward a “common definition of English learner”: Guidance for states and state assessment consortia in defining and addressing policy and technical issues and options*. Council of Chief State School Officers. <https://www.wested.org/resources/toward-a-common-definition-of-english-learner-guidance-for-states-and-state-assessment-consortia-in-defining-and-addressing-policy-and-technical-issues-and-options/>
- Linquanti, R., & Cook, G. (2015). *Re-examining reclassification: Guidance from a national working session on policies and practices for exiting students from English learner status*. Council of Chief State School Officers. <https://www.wested.org/resources/toward-a-common-definition-of-english-learner-guidance-for-states-and-state-assessment-consortia-in-defining-and-addressing-policy-and-technical-issues-and-options/>

- Linquanti, R., Cook, G., Bailey, A., & MacDonald, R. (2016). *Moving toward a more common definition of English learner: Collected guidance for states and multi-state assessment consortia*. Council of Chief State School Officers.  
<https://www.wested.org/resources/toward-a-common-definition-of-english-learner-guidance-for-states-and-state-assessment-consortia-in-defining-and-addressing-policy-and-technical-issues-and-options/>
- Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford Press.
- Luo, R., Pace, A., Levine, D., Iglesias, A., de Villiers, J., Golinkoff, R. M., ... Hirsh-Pasek, K. (2021). Home literacy environment and existing knowledge mediate the link between socioeconomic status and language learning skills in dual language learners. *Early Childhood Research Quarterly*, 55, 1–14. <https://doi.org/10.1016/j.ecresq.2020.10.007>
- Macswan, J., Thompson, M. S., Rolstad, K., McAlister, K., & Lobo, G. (2017). Three theories of the effects of language education programs: An empirical evaluation of bilingual and English-only policies. *Annual Review of Applied Linguistics*, 37, 218–240.  
<https://doi.org/10.1017/S0267190517000137>
- Mancilla-Martinez, J., & Lesaux, N. K. (2011). Early home language use and later vocabulary development. *Journal of Educational Psychology*, 103(3), 535–546.  
<https://doi.org/10.1037/a0023655>
- Mavrogordato, M., & White, R. S. (2017). Reclassification variation: How policy implementation guides the process of exiting students from English learner status. *Educational Evaluation and Policy Analysis*, 39(2), 281–310.  
<https://doi.org/10.3102/0162373716687075>
- McField, G. P., & McField, D. R. (2014). The consistent outcome of bilingual education programs: A meta-analysis of meta-analyses. In G. McField (Ed.), *The miseducation of English learners: A tale of three states and lessons to be learned* (pp. 267–297). Information Age Publishing.
- Michigan Department of Education (MDE). (2020). *English learner program entrance and exit protocol*.  
[https://www.michigan.gov/documents/mde/EL\\_Entrance\\_Exit\\_Protocol\\_722328\\_7.pdf](https://www.michigan.gov/documents/mde/EL_Entrance_Exit_Protocol_722328_7.pdf)
- Menken, K. (2006). Teaching to the test: How No Child Left Behind impacts language policy, curriculum, and instruction for English language learners. *Bilingual Research Journal*, 30(2), 521–546. <https://doi.org/10.1080/15235882.2006.10162888>
- Menken, K. (2008). *English Learners Left Behind: Standardized Testing as Language Policy*. Multilingual Matters.
- Menken, K., & Kleyn, T. (2010). The long-term impact of subtractive schooling in the educational experiences of secondary English language learners. *International Journal of*

- Bilingual Education and Bilingualism*, 13(4), 399–417.  
<https://doi.org/10.1080/13670050903370143>
- Menken, K., Kleyn, T., & Chae, N. (2012). Spotlight on “long-term English language learners”: Characteristics and prior schooling experiences of an invisible population. *International Multilingual Research Journal*, 6(2), 121–142.  
<https://doi.org/10.1080/19313152.2012.665822>
- Menken, K., & Solorza, C. (2014). No Child Left Bilingual: Accountability and the elimination of bilingual education programs in New York City schools. *Educational Policy*, 28(1), 96–125. <https://doi.org/10.1177/0895904812468228>
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256. <https://doi.org/10.1177/026553229601300302>
- Micheltmore, K., & Dynarski, S. (2017). The gap within the gap: Using longitudinal data to understand income differences in educational outcomes. *AERA Open*, 3(1), 1–18.  
<https://doi.org/10.1177/2332858417692958>
- Nagy, B. (2004). Calculating distance with neighborhood sequences in the hexagonal grid. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3322(2), 98–109.  
[https://doi.org/10.1007/978-3-540-30503-3\\_8](https://doi.org/10.1007/978-3-540-30503-3_8)
- Nakamoto, J., Lindsey, K. A., & Manis, F. R. (2007). A longitudinal analysis of English language learners’ word decoding and reading comprehension. *Reading and Writing*, 20(7), 691–719. <https://doi.org/10.1007/s11145-006-9045-7>
- National Academies of Sciences, Engineering, and Medicine. (2017). *Promoting the educational success of children and youth learning English: Promising futures*. The National Academies Press. <https://doi.org/10.17226/24677>
- National Center for Education Statistics. (2017). *NAEP report card: Mathematics*.  
<https://www.nationsreportcard.gov/mathematics/?grade=4>
- National Center for Education Statistics. (2017). *NAEP report card: Reading*.  
<https://www.nationsreportcard.gov/reading/?grade=4>
- National Center for Education Statistics. (2019). *Digest of education statistics*.  
[https://nces.ed.gov/programs/digest/d19/tables/dt19\\_204.27.asp?current=yes](https://nces.ed.gov/programs/digest/d19/tables/dt19_204.27.asp?current=yes)
- No Child Left Behind Act, 20 U.S.C. § 6319 (2002).  
<https://www2.ed.gov/admins/lead/account/nclbreference/reference.pdf>
- O’Grady, W. (2008). The emergentist program. *Lingua*, 118(4), 447–464.  
<https://doi.org/10.1016/j.lingua.2006.12.001>

- Office of English Language Acquisition. (2021). *Profile of English learners in the United States*. [https://ncela.ed.gov/sites/default/files/fast\\_facts/DEL4.4\\_ELProfile\\_508\\_1.4.2021\\_OELA.pdf](https://ncela.ed.gov/sites/default/files/fast_facts/DEL4.4_ELProfile_508_1.4.2021_OELA.pdf)
- Office of English Language Acquisition. (2020). *English learners with disabilities*. [https://ncela.ed.gov/sites/default/files/fast\\_facts/DEL4.4\\_ELProfile\\_508\\_1.4.2021\\_OELA.pdf](https://ncela.ed.gov/sites/default/files/fast_facts/DEL4.4_ELProfile_508_1.4.2021_OELA.pdf)
- Olsen, L. (2014). *Meeting the unique needs of long term English language learners: A guide for educators*. National Education Association. <https://rcoe.learning.powerschool.com>
- Olson, C. B., Scarcella, R., & Matuchniak, T. (2015). English learners, writing, and the common core. *The Elementary School Journal*, 162(4), 570–592. <https://doi.org/10.1086/681235>
- Ou, S. R., & Reynolds, A. J. (2010). Grade retention, postsecondary education, and public aid receipt. *Educational Evaluation and Policy Analysis*, 32(1), 118–139. <https://doi.org/10.3102/0162373709354334>
- Paradis, J. (2011). Individual differences in child English second language acquisition. *Linguistic Approaches to Bilingualism*, 1(3), 213–237. <https://doi.org/10.1075/lab.1.3.01par>
- Paradis, J. (2008). Second Language Acquisition in Childhood. In E. Hoff & M. Shatz (Eds.), *Blackwell Handbook of Language Development* (pp. 387–405). Blackwell. <https://doi.org/10.1002/9780470757833.ch19>
- Paradis, J., Rusk, B., Sorenson Duncan, T., & Govindarajan, K. (2017). Children’s second language acquisition of English complex syntax: The role of age, input, and cognitive factors. *Annual Review of Applied Linguistics*, 37, 148–167. <https://doi.org/10.1017/S0267190517000022>
- Paradis, J., Soto-Corominas, A., Chen, X., & Gottardo, A. (2020). How language environment, age, and cognitive capacity support the bilingual development of Syrian refugee children recently arrived in Canada. *Applied Psycholinguistics*, 41(6), 1255–1281. <https://doi.org/10.1017/S014271642000017X>
- Pearson, B. Z. (2007). Social factors in childhood bilingualism in the United States. *Applied Psycholinguistics*, 28(3), 399–410. <https://doi.org/10.1017/S014271640707021X>
- Perin, D., De La Paz, S., Piantedosi, K. W., & Peercy, M. M. (2017). The writing of language minority students: A literature review on its relation to oral proficiency. *Reading and Writing Quarterly*, 33(5), 465–483. <https://doi.org/10.1080/10573569.2016.1247399>
- Planty, M., Hussar, W., Snyder, T., Kena, G., Kewalramani, A., Kemp, J., ... Dinkes, R. (2009). *The condition of education 2009*. U.S. Department of Education, Institute of Education Sciences. <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009081>



- Quiroz, B. G., Snow, C. E., & Zhao, J. (2010). Vocabulary skills of Spanish-English bilinguals: Impact of mother-child language interactions and home language and literacy support. *International Journal of Bilingualism*, 14(4), 379–399. <https://doi.org/10.1177/1367006910370919>
- Reese, L., Garnier, H., & Gallimore, R. (2000). Longitudinal analysis of the antecedents of emergent Spanish literacy and middle-school English reading achievement of Spanish-speaking students. *American Educational Research Journal*, 37(3), 633–662. <https://doi.org/doi.org/10.2307/1163484>
- Robinson-Cimpian, J. P., Thompson, K. D., & Umansky, I. M. (2016). Research and policy considerations for English learner equity. *Policy Insights from the Behavioral and Brain Sciences*, 3(1), 129–137. <https://doi.org/10.1177/2372732215623553>
- Roderick, M., & Nagaoka, J. (2005). Retention under Chicago’s high-stakes testing program: Helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis*, 27(4), 309–340. <https://www.jstor.org/stable/3699564>
- Rolstad, K., Mahoney, K. S., & Glass, G. V. (2005). Weighing the evidence: A meta-analysis of bilingual education in Arizona. *Bilingual Research Journal*, 29(1), 43–67. <https://doi.org/10.1080/15235882.2005.10162823>
- Rossell, C., & Baker, K. (1996). The educational effectiveness of bilingual education. *Research in the Teaching of English*, 30(1), 7–74. <https://www.jstor.org/stable/40171543%0AJSTOR>
- Saunders, W. M., & Marcelletti, D. J. (2013). The gap that can’t go away: The catch-22 of reclassification in monitoring the progress of English learners. *Educational Evaluation and Policy Analysis*, 35(2), 139–156. <https://doi.org/10.3102/0162373712461849>
- Savage, R., Kozakewich, M., Genesee, F., Erdos, C., & Haigh, C. (2017). Predicting writing development in dual language instructional contexts: exploring cross-linguistic relationships. *Developmental Science*, 20(1). <https://doi.org/10.1111/desc.12406>
- Schissel, J. L., & Kangas, S. E. N. (2018). Reclassification of emergent bilinguals with disabilities: the intersectionality of improbabilities. *Language Policy*, 17(4), 567–589. <https://doi.org/10.1007/s10993-018-9476-4>
- Schwerdt, G., West, M. R., & Winters, M. A. (2017). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. *Journal of Public Economics*, 152, 154–169. <https://doi.org/10.1016/j.jpubeco.2017.06.004>
- Singer, J. D., & Willet, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

- Slama, R. B. (2014). Investigating whether and when English learners are reclassified into mainstream classrooms in the United States: A discrete-time survival analysis. *American Educational Research Journal*, 51(2), 220–252. <https://doi.org/10.3102/0002831214528277>
- Slama, R. B. (2012). A longitudinal analysis of academic English proficiency outcomes for adolescent English language learners in the United States. *Journal of Educational Psychology*, 104(2), 265–285. <https://doi.org/10.1037/a0025861>
- Slama, R., Molefe, A., Gerdeman, D., Herrera, A., Brodziak de los Reyes, I., August, D., & Cavazos, L. (2017). *Time to proficiency for Hispanic English learner students in Texas. (REL 2018-280)*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. <https://login.proxy.lib.fsu.edu/login?url=https://search.proquest.com/docview/2009554575?accountid=4840>
- Slavin, R. E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research*, 75(2), 247–284. <https://doi.org/10.3102/00346543075002247>
- Slavin, R. E., Madden, N., Calderón, M., Chamberlain, A., & Hennessy, M. (2011). Reading and language outcomes of a multiyear randomized evaluation of transitional bilingual education. *Educational Evaluation and Policy Analysis*, 33(1), 47–58. <https://doi.org/10.3102/0162373711398127>
- Steele, J. L., Slater, R. O., Zamarro, G., Miller, T., Li, J., Burkhauser, S., & Bacon, M. (2017). Effects of dual-language immersion programs on student achievement: Evidence from lottery data. *American Educational Research Journal*, 54(S1), 282S-306S. <https://doi.org/10.3102/0002831216634463>
- Stevens, Paul, B. (2006). Is Spanish really so easy? Is Arabic really so hard?: Perceived difficulty in learning Arabic as a second language. In K. M. Wahba, Z. A. Taha, & L. England (Eds.), *Handbook for Arabic Language Teaching Professionals in the 21st Century* (pp. 35–63). Lawrence Erlbaum Associates.
- Sugarman, J., & Geary, C. (2018). *English learners in Michigan: Demographics, outcomes, and state accountability policies*. Migration Policy Institute. [https://www.migrationpolicy.org/sites/default/files/publications/EL-factsheet2018-Michigan\\_Final.pdf](https://www.migrationpolicy.org/sites/default/files/publications/EL-factsheet2018-Michigan_Final.pdf)
- Surrain, S. (2018). ‘Spanish at home, English at school’: How perceptions of bilingualism shape family language policies among Spanish-speaking parents of preschoolers. *International Journal of Bilingual Education and Bilingualism*. <https://doi.org/10.1080/13670050.2018.1546666>



- Tavassolie, T., & Winsler, A. (2019). Predictors of mandatory 3rd grade retention from high-stakes test performance for low-income, ethnically diverse children. *Early Childhood Research Quarterly*, 48, 62–74. <https://doi.org/10.1016/j.ecresq.2019.02.002>
- Thompson, K. D. (2017). English learners' time to reclassification: An analysis. *Educational Policy*, 31(3), 330–363. <https://doi.org/10.1177/0895904815598394>
- Tomasello, M. (2003). *Constructing a language*. Harvard University Press.
- U.S. Department of Education. (2016). *Non-regulatory guidance: English learner and Title III of the Elementary and Secondary Education Act (ESEA), as amended by the Every Student Succeeds Act (ESSA)*. <https://www2.ed.gov/policy/elsec/leg/essa/essatitleiiiiguidenglishlearners92016.pdf>
- U.S. Department of Education, & Office of English Language Acquisition. (2017). *English learner toolkit for state and local education agencies (SEAs and LEAs)*. [https://ncela.ed.gov/files/english\\_learner\\_toolkit/OELA\\_2017\\_ELsToolkit\\_508C.pdf](https://ncela.ed.gov/files/english_learner_toolkit/OELA_2017_ELsToolkit_508C.pdf)
- Umansky, I. M., Callahan, R. M., & Lee, J. C. (2020). Making the invisible visible: Identifying and interrogating ethnic differences in English learner reclassification. *American Journal of Education*, 126(3), 335–388. <https://doi.org/10.1086/708250>
- Umansky, I. M., & Reardon, S. F. (2014). Reclassification patterns among Latino English learner students in bilingual, dual immersion, and English immersion classrooms. *American Educational Research Journal*, 51(5), 879–912. <https://doi.org/10.3102/0002831214545110>
- Umansky, I. M., Reardon, S. F., Hakuta, K., Thompson, K. D., Estrada, P., Hayes, K., ... Goldenberg, C. (2015). Improving the opportunities and outcomes of California's students learning English: Findings from school district-university collaborative partnerships. *PACE Policy Brief*, 15(1). [http://edpolicyinca.org/sites/default/files/PACE Policy Brief 15-1\\_v6.pdf](http://edpolicyinca.org/sites/default/files/PACE%20Policy%20Brief%2015-1_v6.pdf)
- Umansky, I. M., Thompson, K. D., & Díaz, G. (2017). Using an ever-English learner framework to examine disproportionality in special education. *Exceptional Children*, 84(1), 76–96. <https://doi.org/10.1177/0014402917707470>
- Valbuena, J., Mediavilla, M., Choi, Á., & Gil, M. (2021). Effects of grade retention policies: A literature review of empirical studies applying causal inference. *Journal of Economic Surveys*, 35(2), 408–451. <https://doi.org/10.1111/joes.12406>
- WIDA. (2020). *WIDA English Language Development Standards Framework, 2020 edition: Kindergarten—Grade 12*. Wisconsin Center for Education Research. <https://wida.wisc.edu/sites/default/files/resource/WIDA-ELD-Standards-Framework-2020.pdf>

- Wiley, T. G., & Wright, W. E. (2004). Against the undertow: Language-minority education policy and politics in the “age of accountability.” *Educational Policy*, 18(1), 142–168. <https://doi.org/10.1177/0895904803260030>
- Williams, C., & Lowrance-Faulhaber, E. (2018). Writing in young bilingual children: Review of research. *Journal of Second Language Writing*, 42, 58–69. <https://doi.org/10.1016/j.jslw.2018.10.012>
- Winke, P. (2011). Evaluating the validity of a high-stakes ESL test: Why teachers’ perceptions matter. *TESOL Quarterly*, 45(4), 628–660. <https://doi.org/10.5054/tq.2011.268063>
- Winke, P., & Zhang, X. (2019). How a third-grade reading retention law will affect ELLs in Michigan, and a call for research on child ELL reading development. *TESOL Quarterly*, 53(2), 529–542. <https://doi.org/10.1002/tesq.481>
- Winters, M. A., & Greene, J. P. (2012). The medium-run effects of Florida’s test-based promotion policy. *Education Finance and Policy*, 7(3), 305–330. [https://doi.org/10.1162/EDFP\\_a\\_00069](https://doi.org/10.1162/EDFP_a_00069)
- Wolf, M. K., Kao, J., Griffin, N., Herman, J. L., Bachman, P. L., Chang, S. M., & Farnsworth, T. (2008). *Issues in assessing English language learners: English language proficiency measures and accommodation uses—Practice review (CSE Tech. Rep. No. 732)*. Los Angeles, CA: University of California, National Center for Research Evaluation, Standards, and Student Testing (CRESST). <https://files.eric.ed.gov/fulltext/ED502284.pdf>
- Wong Fillmore, L. (1991). When learning a second language means losing the first. *Early Childhood Research Quarterly*, 6, 323–346. [https://doi.org/10.1016/S0885-2006\(05\)80059-6](https://doi.org/10.1016/S0885-2006(05)80059-6)
- Zehler, A. M., Fleischman, H. L., Hopstock, P. J., Pendzick, M. L., & Stephenson, T. G. (2003). *Descriptive study of services to LEP students and LEP students with disabilities: Findings on special education LEP students*. U.S. Department of Education, Office of English Language Acquisition.
- Zhang, X., Winke, P., & Clark, S. (2020). Background characteristics and oral proficiency development over time in lower-division college foreign language programs. *Language Learning*, 70(3), 807–847. <https://doi.org/10.1111/lang.12396>